

**THE USE OF WATER FEATURES FOR IMPROVING
SPEECH PRIVACY AND COGNITIVE PERFORMANCE
IN OPEN-PLAN OFFICES**

Zanyar Anwer Abdalrahman

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Energy, Geoscience, Infrastructure and Society

March 2018

The copyright in this thesis is owned by the author. Any quotation from the thesis or use of any of the information contained in it must acknowledge this thesis as the source of the quotation or information.

ABSTRACT

Noise in open-plan offices, especially irrelevant speech, has detrimental effects on the psychological and cognitive functioning of people. Noise masking has proven to be effective in reducing these effects. Within that context, this research examines the use of water features in masking irrelevant speech and improving the sound environment in open-plan offices. The research comprised five experiments.

Experiment 1 examined the preferred sound pressure level of water sounds when used to mask irrelevant speech. The preferred sound pressure level of water sounds was found to be 45 dBA. In addition, the preferred masking level was independent from the intelligibility level of irrelevant speech as well as the type of the water sounds. Experiment 2 examined audio-only and audio-visual preferences of six water sounds when used to mask irrelevant speech. A 37-jet fountain was preferred in the audio-only condition, whilst a 4-step cascade was preferred in the audio-visual condition. The audio-visual condition increased the likelihood of making positive changes in people's perception by 1.1 to 2.5 times, in comparison to the audio-only condition.

Experiment 3 examined the effect of masking irrelevant speech on cognitive performance and subjective satisfaction. Results showed that cognitive performance of a serial recall task, as well as subjective satisfaction, to be significantly higher when irrelevant speech was masked by a water sound, in comparison to a speech-only condition. The gender of participants was found to have a significant effect on cognitive performance. Female participants' performance was lower than their male counterparts. Furthermore, female participants benefited more from the masking of irrelevant speech.

Experiment 4 examined the longer-term effects of adding a water feature in an open-plan office. Satisfaction with the sound environment significantly increased after installing a water feature in an open-plan office. Experiment 5 measured the reduction in the distraction distance associated with installing a water feature in two open-plan offices. The reduction in the distraction distance was measured to be between 8.64 m and 10.05 m, depending on the space tested. The layout of the workspaces played a key role in dictating the importance of the reduction in the distraction distance.

Following recommendations and design criteria given in this study, it is possible to use water features in open-plan offices as a means of masking irrelevant speech and creating a pleasant soundscape and work environment which promote cognitive performance and speech privacy and increase subjective satisfaction.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr Laurent Galbrun, who made it possible for me to complete this project, through his competence and immense knowledge of the field. I feel fortunate and privileged to have had Dr Galbrun as my supervisor. He has been a true mentor, teacher and a dear friend to me. His excellent and spot on feedback, alongside his positive and forward-thinking attitude, have significantly enhanced the quality of this research.

I would like to extend my gratitude to Dr Jennifer A. Veitch from the National Research Council Canada, Institute for Research in Construction (IRC), who very kindly agreed to provide the speech recording that was used throughout this project. Dr Veitch's contribution to this study is truly appreciated.

Many thanks go to both Alastair MacFarlane and David Murray for giving me a helping hand in modifying the water feature. I would also like to thank the people who participated in this study for their genuine response, and the staff members of the research support office, especially Anne Ormston, for giving their kind permission to install a water feature in their office. I also wish to acknowledge the generosity of the School of EGIS for providing financial support to this research through the James Watt Scholarship.

I am grateful to my two little nieces, Anya and Avia for making my journey full of joy, despite the distance between us, through their constant video calls. I thank my brothers, Karzan, Didar and Divar, and my sisters, Avan and Sava for believing in me and providing me with unconditional support to complete this project as best as I could. A truly special thank goes to the two angels of my life, my mother and my father, who simply through their existence, made my life happy. I could have never even managed to start this journey, let alone complete it, without their never-ending and unstoppable love, guidance, and support.

Last but not least, I would like to thank my lovely wife, Hazhan Majeed, to whom I dedicate this thesis. You have been a great source of inspiration for me. Your continuous and unquestionable support is beyond compare, for that I am forever indebted. You provided me with the stamina to continue this journey and realise my dreams. Your existence in my life gives me hope and turns any sorrow to ecstasy. I am extremely lucky to have you in my life. You have been, and always remain the true love of my life.

Research Thesis Submission

Name:	Zanyar Anwer Abdalrahman		
School:	School of Energy, Geoscience, Infrastructure and Society		
Version: <i>(i.e. First, Resubmission, Final)</i>	Final	Degree Sought:	PhD Construction

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

- 1) the thesis embodies the results of my own work and has been composed by myself
- 2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
- 3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
- 4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
- 5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
- 6) I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

* Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.

Signature of Candidate:		Date:	26/03/2018
-------------------------	--	-------	------------

Submission

Submitted By <i>(name in capitals)</i> :	ZANYAR ABDALRAHMAN
Signature of Individual Submitting:	
Date Submitted:	26/03/2018

For Completion in the Student Service Centre (SSC)

Received in the SSC by <i>(name in capitals)</i> :			
Method of Submission <i>(Handed in to SSC; posted through internal/external mail):</i>			
e-thesis Submitted <i>(mandatory for final theses)</i>			
Signature:		Date:	

Please note this form should be bound into the submitted thesis.

TABLE OF CONTENTS

LIST OF FIGURES	xii
FIGURES CREDITS	xix
LIST OF TABLES	xxiv
GLOSSARY OF SYMBOLS.....	xxvi
LIST OF CONFERENCES AND PUBLICATIONS	xxx
CHAPTER 1: INTRODUCTION	1
1.1 General introduction.....	1
1.2 Justification of the research.....	3
1.3 Research aim and objectives	4
1.4 Methodology	5
1.4.1 Experiment 1: Sound level preference test.....	5
1.4.2 Experiment 2: Audio-only and audio-visual preference and perception tests.....	5
1.4.3 Experiment 3: Cognitive performance and subjective satisfaction.....	6
1.4.4 Experiment 4 Longer-term exposure to a water sound	6
1.4.5 Experiment 5 Objective measures.....	6
1.5 Outline of thesis.....	7
CHAPTER 2: LITERATURE REVIEW	9
2.1 Introduction	9
2.2 Psychoacoustics.....	9
2.2.1 Pitch and pitch strength.....	10
2.2.2 Loudness	10
2.2.3 Sharpness	10
2.2.4 Roughness and fluctuation strength	10
2.2.5 Weighting filters.....	11
2.3 Exposure to natural environments	12

2.4	Water features.....	13
2.4.1	Water sound generation.....	16
2.4.2	Water features and Architecture.....	17
2.4.3	The indispensable water.....	18
2.5	A different approach to noise abatement.....	19
2.5.1	The soundscape approach	20
2.5.2	Soundscape assessment.....	23
2.6	Sound masking	23
2.6.1	Acoustic space.....	24
2.7	Human speech	25
2.8	Speech intelligibility.....	26
2.8.1	Modulation transfer function	27
2.8.2	Speech Transmission Index (STI).....	29
2.8.3	Factors influencing speech intelligibility.....	30
2.9	Speech masking in open-plan offices	31
2.9.1	Types of speech-masking sounds.....	32
2.9.2	Recommended masking spectrum and level.....	37
2.10	Effect of speech intelligibility on cognitive performance and satisfaction ..	37
2.10.1	Performance decrement as a function of the STI.....	41
2.10.2	Theories behind performance decrement as a result of intelligible speech.....	43
2.11	Preference of water sounds	45
2.12	Audio-visual preferences of water features	56
2.13	Single-number quantities	59
2.14	Discussion.....	61
2.15	Conclusions.....	65
CHAPTER 3: METHODOLOGY AND STATISTICAL ANALYSIS.....		72
3.1	Introduction	72

3.2	Water sounds	72
3.2.1	Test structure	73
3.3	Speech recording	75
3.4	Methodology	76
3.4.1	Experiment 1: Sound level preference test.....	76
3.4.2	Experiment 2: Audio-only and audio-visual preference and perception tests.....	77
3.4.3	Experiment 3: Cognitive performance and subjective satisfaction.....	78
3.4.4	Experiment 4 Longer-term exposure to a water sound	78
3.4.5	Experiment 5 Objective measures.....	79
3.5	Statistical analysis	79
3.5.1	The mean.....	80
3.5.2	Confidence intervals.....	80
3.5.3	Bootstrap Confidence intervals	80
3.5.4	Outliers.....	81
3.5.5	Null hypothesis significance testing and the <i>p</i> -value.....	81
3.5.6	Comparing two means.....	81
3.5.7	Comparing several means	82
3.5.8	Factorial designs.....	83
3.5.9	Post hoc procedure	83
3.5.10	Correlation analysis.....	84
3.5.11	Effect size.....	84
3.5.12	Reliability analysis.....	85
3.5.13	Categorical analysis (chi-square test).....	85
3.6	Conclusions	85
CHAPTER 4: SOUND LEVEL PREFERENCE TEST.....		86
4.1	Introduction	86
4.2	Water sounds	86

4.3	Speech recording	87
4.4	Potential masking levels	88
4.5	Intelligibility of background speech.....	88
4.6	Participants	90
4.7	Methodology	91
4.8	Procedure.....	92
4.9	Consistency test.....	93
4.10	Equipment and digital audio processing.....	94
4.11	Statistical analysis.....	94
4.12	Results.....	95
4.12.1	Effect of speech intelligibility and type of water sound on preference scores	96
4.12.2	Preference of sound pressure level of the water sounds	97
4.13	Discussion.....	98
4.14	Conclusions.....	98
CHAPTER 5: AUDIO-ONLY AND AUDIO-VISUAL PREFERENCES AND PERCEPTION		99
5.1	Introduction	99
5.2	Water sounds	99
5.3	Speech recording	100
5.4	Participants	101
5.5	Methodology	102
5.6	Procedure.....	103
5.6.1	Preference test	104
5.6.2	Perception test	106
5.6.3	Consistency test.....	107
5.7	Equipment and digital audio and visual processing	107
5.8	Statistical analysis	108
5.9	Results	109

5.9.1	Audio-only and audio-visual preference results	109
5.9.2	Audio-only and audio-visual perception results	111
5.9.3	Comparison between audio-only and audio-visual perception scores. ...	113
5.9.4	Comparison between preference and perception scores	114
5.9.5	Categorical analysis of perception scores	116
5.9.6	Cluster analysis	118
5.9.7	Correlation between preference/perception ratings and acoustic/psychoacoustic measures	123
5.10	Discussion.....	126
5.11	Conclusions.....	130
CHAPTER 6: COGNITIVE PERFORMANCE AND SUBJECTIVE SATISFACTION.....		133
6.1	Introduction	133
6.2	Selected water sound	133
6.3	Speech recording	134
6.4	Participants	136
6.5	The open-plan space	137
6.6	Background noise conditions	140
6.7	Cognitive tasks	140
6.7.1	Serial recall task	145
6.7.2	One-back task.....	145
6.7.3	Information matching task	146
6.7.4	Reading comprehension task.....	147
6.8	Questionnaires	149
6.8.1	Subjective satisfaction questionnaire	150
6.9	Procedure.....	151
6.10	Equipment and software	152
6.11	Statistical analysis.....	153
6.11.1	Multivariate approach	153

6.11.2	Univariate approach	154
6.11.3	Checking assumptions.....	155
6.11.4	Non-parametric tests	156
6.11.5	Effect size.....	156
6.11.6	Normalisation of scores.....	157
6.12	Results.....	157
6.12.1	Serial recall task	157
6.12.2	Information matching task	165
6.12.3	One-back task.....	169
6.12.4	Reading comprehension task.....	173
6.12.5	Subjective satisfaction.....	181
6.12.6	Subjective task difficulty and speech intelligibility	182
6.12.7	Correlation analysis.....	184
6.13	Discussion.....	186
6.14	Conclusions.....	190
CHAPTER 7: LONGER-TERM EXPOSURE TO A WATER SOUND.....		194
7.1	Introduction	194
7.2	Design of the water feature	194
7.3	Questionnaire.....	198
7.4	Participants	200
7.5	Statistical analysis	200
7.6	Procedure.....	201
7.7	Results	201
7.7.1	Participants' sensitivity to noise.....	202
7.7.2	Satisfaction with the physical work environment	202
7.7.3	Satisfaction with the noise environment	204
7.7.4	Perception of sound environment.....	207
7.7.5	Correlational analysis.....	210

7.8	Discussion	213
7.9	Conclusions	215
CHAPTER 8: OBJECTIVE MEASURES		217
8.1	Introduction	217
8.2	Single-number quantities.....	217
8.3	Water sound.....	218
8.4	Equipment and setup	218
8.5	Measurement procedure	220
8.5.1	Determination of the sound power level of an omnidirectional speaker.....	221
8.5.2	Determination of the spatial decay.....	223
8.5.3	Determination of distraction distance, r_D and privacy distance, r_P	226
8.6	Open-plan offices	228
8.7	Results	232
8.7.1	Office 1.....	232
8.7.2	Office 2.....	235
8.8	Discussion	238
8.9	Conclusions	243
CHAPTER 9: CONCLUSIONS.....		244
9.1	Introduction	244
9.2	Findings	244
9.3	Conclusions	245
9.3.1	Conclusions of the chapters	246
9.4	Impact of the research	248
9.5	Limitations and future work	249
REFERENCES		253
Appendix A: Evaluation form used in the level preference test		263
Appendix B: Instructions given to participants in the level preference test		265

Appendix C: Evaluation form used in the audio-only and audio-visual preference and perception tests	267
Appendix D: Instructions given to participants the audio-only and audio-visual preference and perception tests	270
Appendix E: General instructions for the cognitive performance tasks	275
Appendix F: Serial recall task and instructions.....	279
Appendix G: One-back task and instructions.....	280
Appendix H: Information matching task and instructions	281
Appendix I: Reading comprehension task and instructions	282
Appendix J: Satisfaction questionnaire used in Experiment 4 (Part 1).....	285
Appendix K: Satisfaction questionnaire used in Experiment 4 (Part2)	290
Appendix L: Participants' response to satisfaction questionnaire in Experiment 4.....	292

LIST OF FIGURES

Figure 2.1 A-weighting and C-weighting curves.	11
Figure 2.2 Examples of water features in nature.	13
Figure 2.3 Different water patterns of fountains.	14
Figure 2.4 A waterfall in a built environment.	15
Figure 2.5 A multiple-step cascade in Mount Baranduda, Australia.	15
Figure 2.6 Two examples of indoor water features.	16
Figure 2.7 Alhambra pool and fountains in Granada, Spain.	17
Figure 2.8 Fallingwater House by Frank Lloyd Wright.	18
Figure 2.9 Elements in the perceptual construct of soundscape.	21
Figure 2.10 Factors to be considered in indoor soundscape studies.	22
Figure 2.11 Acoustic spaces before and after installing a masking system.	24
Figure 2.12 The audible spectrum, and the modulation spectrum of human speech.	26
Figure 2.13 The Modulation Transfer Function (MTF).	28
Figure 2.14 Mean error rates in the serial recall task under seven sound conditions.	33
Figure 2.15 Acoustic satisfaction under seven sound conditions.	34
Figure 2.16 One-third octave band frequency spectra of four water sounds and a filtered pink noise used in Hongisto <i>et al.</i> (2017).	36
Figure 2.17 The schematic prediction model, showing the decrease in performance as a function of the STI, irrespective of the sound level of speech.	41
Figure 2.18 Task-averaged performance change between acoustic conditions as a function of the STI alongside Hongisto's model (2005).	42
Figure 2.19 Subjective speech intelligibility as a function of the STI for sentences (non-optimized SRT) from which Hongisto's model (2005) was derived.	43
Figure 2.20 Normalised preference levels of natural sounds used to mask two urban noises, road traffic noise, and construction noise.	47
Figure 2.21 Visual images of water features and a street used by Jeon <i>et al.</i> (2012). ...	48
Figure 2.22 Spectra produced by different water flow rates.	50
Figure 2.23 Normalised preference levels of water sounds examined by Galbrun and Ali (2013).	52

Figure 2.24 Semantic characterization of each water sound used over road traffic noise, illustrating both attributes and components.	54
Figure 2.25 Audio-only and audio-visual preference scores of water sounds combined with road traffic noise at an L_{Aeq} of 55 dBA.....	57
Figure 2.26 Preferred water features from audio-only, visual-only, and audio-visual tests: normalised preference values as a function of water features.....	58
Figure 3.1 Laboratory structure used to generate the water sounds.....	74
Figure 3.2 Photographs of water features selected in this study, taken at the Building Services laboratory of Heriot-Watt University.	74
Figure 4.1 Octave-band spectra of irrelevant speech (48 dBA), CA (48 dBA), FTW (48 dBA), and background noise (43 dBA).....	89
Figure 4.2 Age distribution of participants in the sound level preference test.....	90
Figure 4.3 Normalised preference scores for the four test conditions of the sound level preference test, alongside the averaged preference scores.....	95
Figure 4.4 Number of times the "no preference" option was chosen for each pair of sounds.	96
Figure 5.1 Octave-band spectra of six water sounds (45 dBA) and the spectrum of irrelevant speech (48 dBA) used in the audio-only and audio-visual tests.	100
Figure 5.2 Age distribution of participants in audio-only and audio-visual preference and perception tests.	102
Figure 5.3 Still images from animations of the water features used in the audio-visual condition of the preference and perception tests.	105
Figure 5.4 Normalised audio-only and audio-visual preference scores for six water sounds used in this study	109
Figure 5.5 Change in participants' perception caused by six waterscapes in audio-only and audio-visual conditions.	111
Figure 5.6 Normalised preference and perception scores given in the audio-only condition.....	114
Figure 5.7 Normalised preference and perception scores given in the audio-visual condition.....	115
Figure 5.8 Normalised audio-only and audio-visual preference scores in Cluster 1 (17 participants).....	119
Figure 5.9 Normalised audio-only and audio-visual preference scores in Cluster 2 (14 participants).....	120

Figure 5.10 Normalised audio-only and audio-visual perception scores in Cluster 1 (17 participants).	122
Figure 5.11 Normalised audio-only and audio-visual perception scores in Cluster 2 (14 participants).	123
Figure 5.12 Octave-band spectra of speech (48 dBA) and four highly preferred water sounds in the preference test, against the reference masking spectrum recommended by Veitch <i>et al.</i> (2002), calibrated to 45 dBA.	126
Figure 5.13 Masking curves at 40 dB for two pure tones, 2.4 kHz and 3.5 kHz.	127
Figure 5.14 Spectrogram of two plosives consonantsx, spoken between two vowel sounds.	129
Figure 5.15 Spectrogram of five fricative consonant, spoken between two vowel sounds.	129
Figure 6.1 The spectrum of the 37-jet fountain (FTW) and the energy averaged spectrum of speech (averaged over 17 minutes at 1-second intervals)..	135
Figure 6.2 A-weighted equivalent sound pressure level $L_{Aeq, 5s}$ of the speech recording used in the current study, for the duration of the entire speech recording.....	135
Figure 6.3 Age distribution of participants in the cognitive task performance experiment.	136
Figure 6.4 Test room in which the task performance experiment was carried out..	138
Figure 6.5 A sectional perspective of the test room showing the arrangement of furniture, workstations and speakers.	137
Figure 6.6 A photograph of the test room showing the workstations.	138
Figure 6.7 One-third octave band reverberation time (in seconds) of the test room....	139
Figure 6.8 A screenshot of the interactive array of digits used in the serial recall task.....	145
Figure 6.9 A screenshot of the information matching task.	147
Figure 6.10 Normalised RA and RT (reversed) scores in the serial recall task under the speech-only and the masked speech conditions.	158
Figure 6.11 Averaged normalised performance score in the serial recall task for male and female participants, irrespective of background noise condition.....	159
Figure 6.12 Averaged normalised performance score (RA and RT) in the serial recall task for male and female participants under speech-only and masked speech conditions.	159

Figure 6.13 Normalised RA and RT (reversed) scores in the serial recall task for male and female participants, irrespective of the background noise condition.	160
Figure 6.14 Normalised RA and RT (reversed) scores in the serial recall task under speech-only and masked speech conditions for both male and female participants.	160
Figure 6.15 Global normalised performance score (RA and RT) in the serial recall task for native and non-native English speakers.	161
Figure 6.16 Averaged normalised performance score in the serial recall task for native and non-native English speakers under speech-only and masked speech conditions.	162
Figure 6.17 Normalised RA and RT (reversed) scores in the serial recall task for native and non-native English speakers irrespective of the background noise condition.	162
Figure 6.18 Normalised RA and RT (reversed) scores in the serial recall task for native and non-native English speakers under speech-only and masked speech conditions.	164
Figure 6.19 Relationship between normalised RA and RT (reversed) scores in the serial recall task with NoiseQ scores under speech-only and masked speech conditions.	165
Figure 6.20 Normalised RT score (reversed) in the information masking task, under the speech-only and the masked speech conditions.	166
Figure 6.21 Averaged normalised RT score (reversed) in the information masking task for male and female participants.	167
Figure 6.22 Averaged normalised RT score (reversed) in the information masking task for male and female participants under speech-only and masked speech conditions.	167
Figure 6.23 Averaged normalised RT score (reversed) in the information masking task for native and non-native participants.	168
Figure 6.24 Averaged normalised RT score (reversed) in the information masking task for native and non-native participants under speech-only and masked speech conditions..	168
Figure 6.25 Relationship between normalised RT scores (reversed) in the information masking task with NoiseQ scores under speech-only and masked speech conditions.	169

Figure 6.26 Normalised RT score (reversed) in the one-back task under speech-only and masked speech conditions.	170
Figure 6.27 Averaged normalised RT score (reversed) for male and female participants in the one-back task.	171
Figure 6.28 Averaged normalised RT score (reversed) for male and female participants in the one-back task under speech-only and masked speech conditions.	171
Figure 6.29 Averaged normalised RT score (reversed) for native and non-native participants in the one-back task.	172
Figure 6.30 Averaged normalised RT score (reversed) for native and non-native participants in the one-back task, under speech-only and masked speech conditions..	172
Figure 6.31 Relationship between normalised RT scores (reversed) with NoiseQ scores in the one-back task under speech-only and masked speech conditions.	173
Figure 6.32 Normalised RA and RT (reversed) in the reading comprehension task, under the speech-only condition and the masked speech condition.....	174
Figure 6.33 Averaged normalised performance score for male and female participants in the reading comprehension task..	175
Figure 6.34 Averaged normalised performance score for male and female participants in the reading comprehension task under speech-only and masked speech conditions..	175
Figure 6.35 Normalised RA and RT (reversed) scores task for male and female participants in reading comprehension task, irrespective of background noise condition..	176
Figure 6.36 Normalised RA and RT (reversed) scores in the reading comprehension task, under speech-only and masked speech conditions for both male and female participants.	176
Figure 6.37 Global normalised performance score for native and non-native English speakers in the reading comprehension task..	177
Figure 6.38 Averaged normalised performance score for native and non-native English speakers scores in the reading comprehension task under speech-only and masked speech conditions.	178

Figure 6.39 Normalised RA and RT (reversed) scores for native and non-native English speakers in the reading comprehension task, irrespective of the background noise condition.	178
Figure 6.40 Normalised RA and RT (reversed) scores for native and non-native English speakers in the reading comprehension task under speech-only and masked speech conditions.	179
Figure 6.41 Relationship between normalised RA and RT (reversed) scores in the reading comprehension task with NoiseQ scores under speech-only and masked speech conditions.	180
Figure 6.42 Normalised RA score for the reading comprehension components under speech-only and masked speech conditions.....	181
Figure 6.43 Subjective satisfaction with the overall sound environment and on the individual items' level..	182
Figure 6.44 Subjective task difficulty in speech-only and masked speech conditions.	183
Figure 6.45 Subjective speech intelligibility in speech-only and masked speech condition.....	183
Figure 7.1 Three-step cascade used in this study.	195
Figure 7.2 Directivity (left) and positions at which the sound pressure levels of the water feature was measured (right) in the anechoic chamber.	196
Figure 7.3 Plan of the open-plan office where the water feature was placed in. dimensions are given in metres.	197
Figure 7.4 A photograph of the open-plan office where the water feature was placed in.	197
Figure 7.5 Octave-band spectrum of the background noise with and without the water sound space averaged over 12 workstations.....	198
Figure 7.6 Satisfaction level within Comfort/Functionality aspects of the physical work environment, before and after installing the water feature.	203
Figure 7.7 Satisfaction level within Control/Privacy aspects of the physical work environment, before and after installing the water feature.....	203
Figure 7.8 Satisfaction with noise environment before and after installing the water feature.	205
Figure 7.9 Average ranking scores for 8 noise sources according to their annoyance level.	206
Figure 7.10 Satisfaction level within the sound environment, before and after installing the water feature.	207

Figure 8.1 Equipment used to measure the single-number quantities.....	219
Figure 8.2 A photograph of Office 1.....	229
Figure 8.3 A photograph of Office 2.....	229
Figure 8.4 Plan and dimensions of Office 1.....	230
Figure 8.5 Plan and dimensions of Office 2.....	231
Figure 8.6 Spatial decay rate of the A-weighted sound pressure level of speech in Office 1 from source position 1.	233
Figure 8.7 Spatial decay rate of the A-weighted sound pressure level of speech in Office 1 from source position 2.	234
Figure 8.8 Rate of drop in the STI in Office 1 measured for source position 1, in two masking conditions: no masking and water sound.....	234
Figure 8.9 Rate of drop in the STI in Office 1 measured for source position 2, in two masking conditions: no masking and water sound.....	234
Figure 8.10 Spatial decay rate of the A-weighted sound pressure level of speech in Office 2 from source position 1.....	236
Figure 8.11 Spatial decay rate of the A-weighted sound pressure level of speech in Office 2 from source position 2.....	237
Figure 8.12 Rate of drop in the STI in Office 2 measured for source position 1, in two masking conditions: no masking and water sound.....	237
Figure 8.13 Rate of drop in STI in Office 2 measured for source position 2, in two masking conditions: no masking and water sound.....	237
Figure 8.14 Increase in the annoyance level as a function of the increase in the distraction distance.	238
Figure 8.15 Octave-band spectra of air-conditioning noise (44.3 dBA), water sound (43.9 dBA) and air-conditioning noise and water sound combined (46.6 dBA), measured in Office 2.	239
Figure 8.16 Overlapping of distraction areas of workstations in Office 1.....	241
Figure 8.17 Overlapping of distraction areas of workstations in Office 2.....	242

FIGURES CREDITS

Figure 2.1 A-weighing and C-weighing curves (IEC 61672-1, 2013).

IEC 61672-1 (2013) *Electroacoustics-Sound level meters Part 1: Specifications Measurement microphones*. Brussels: European Committee for Electrotechnical Standardization.

Figure 2.2 Examples of water features in nature.

(a) Natural stream, Pagoeta Natural park, Spain (Flickr: Rafel Miro).

Miro, R. (2017) *Parc Natural de Pagoeta*, Flickr. Available at: <https://goo.gl/ctXK9D> (Accessed: 30 November 2017).

(b) Natural cascade, Mount Baranduda, Australia (Flickr: Dirkus).

Dirkus (2012) *Waterfall, Mount Baranduda*, Flickr. Available at: <https://goo.gl/jonHW8> (Accessed: 30 November 2017).

(c) Niagara Falls, Canada (Flickr: Jess Buttery).

Buttery, J. (2014) *4F5A4571_2_3_4_5_M*, Flickr. Available at: <https://goo.gl/F1io5E> (Accessed: 30 November 2017).

(d) Strokkur geyser in Iceland (Flickr: Ashley Buttle).

Buttle, A. (2006) *Strokkur geyser - geysir*, Flickr. Available at: <https://goo.gl/do5QSb> (Accessed: 30 November 2017).

Figure 2.3 Different water patterns of fountains (Stephen, 1988).

Stephen, S. (1988) 'Water Features: Pools, Cascades, Fountains', *The Architect's Journal*, 2.

Figure 2.4 A waterfall in a built environment (Flickr: ~Pawsitive~Candie_N).

~Pawsitive~Candie_N (2008) *waterfall*, Flickr. Available at: <https://goo.gl/42pM7e> (Accessed: 30 November 2017).

Figure 2.5 A multiple-step cascade in Mount Baranduda, Australia (Flickr: Dirkus).

Dirkus (2012) *Waterfall, Mount Baranduda*, Flickr. Available at: <https://goo.gl/jonHW8> (Accessed: 30 November 2017).

Figure 2.6 Two examples of indoor water features.

(a) Tokyo Disneyland Hotel Lobby's fountain (Flickr: Peter Lee).

Lee, P. (2010) *Tokyo Disneyland Hotel Lobby*, Flickr. Available at: <https://goo.gl/XrRqvG> (Accessed: 30 November 2017).

(b) A small pond-less indoor fountain (Flickr: Jay@MorphoLA).

Jay@MorphoLA (2008) *Weber Residence- Denver, Flickr*. Available at: <https://goo.gl/xhzWfC> (Accessed: 30 November 2017).

Figure 2.7 Alhambra pool and fountains in Granada, Spain (Flickr: John Blower).

Blower, J. (2011) *Granada: Alhambra pool and fountains, Flickr*. Available at: <https://goo.gl/ZeicAk> (Accessed: 30 November 2017).

Figure 2.8 Fallingwater House by Frank Lloyd Wright (Flickr: Nick Stanley).

Stanley, N. (2011) *Fallingwater in the Fall, Flickr*. Available at: <https://goo.gl/wDu8nW> (Accessed: 30 November 2017).

Figure 2.9 Elements in the perceptual construct of soundscape (BS ISO 12913, 2014).

BS ISO 12913 (2014) *Acoustics-Soundscape Part 1 : Definition and conceptual framework*. Geneva: BSI Standards Limited.

Figure 2.10 Factors to be considered in indoor soundscape studies (Dokmeci and Kang, 2010).

Dokmeci, P. N. and Kang, J. (2010) 'Objective parameters for acoustic comfort in enclosed spaces', in *International Congress on Acoustics, ICA 2010*. Sydney, Australia, pp. 1–4.

Figure 2.12 The audible spectrum (a), and the modulation spectrum (b) of human speech (Jacob, 2001).

Jacob, K. (2001) 'Understanding Speech Intelligibility and the Fire Alarm Code', in *National fire Protection Association*. California, USA, May 14.

Figure 2.13 The Modulation Transfer Function (MTF) (Houtgast and Steeneken, 1985).

Houtgast, T. and Steeneken, H. J. M. (1985) 'A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria', *The Journal of the Acoustical Society of America*, 77(3), pp. 1069–1077.

Figure 2.14 Mean error rates (%) in the serial recall task under seven sound conditions (Haapakangas *et al.*, 2011).

Haapakangas, A. *et al.* (2011) 'Effects of five speech masking sounds on performance and acoustic satisfaction. Implications for open-plan offices', *Acta Acustica united with Acustica*, 97, pp. 641–655.

Figure 2.15 Acoustic satisfaction in the seven sound conditions (Haapakangas *et al.*, 2011).

Haapakangas, A. *et al.* (2011) 'Effects of five speech masking sounds on performance and acoustic satisfaction. Implications for open-plan offices', *Acta Acustica united with Acustica*, 97, pp. 641–655.

Figure 2.16 One-third octave band frequency spectra of four water sounds and a filtered pink noise (Hongisto *et al.*, 2017).

Hongisto, V. *et al.* (2017) 'Perception of water-based masking sounds-long-term experiment in an open-plan office', *Frontiers in Psychology*, 8(July) pp.1-14.

Figure 2.17 The schematic prediction model, which shows the decrease in performance as a function of the STI, irrespective of the sound level of speech (Hongisto, 2005).

Hongisto, V. (2005) 'A model predicting the effect of speech of varying intelligibility on work performance', *Indoor Air*, 15(6), pp. 458–468.

Figure 2.18 Task-averaged performance change between acoustic conditions as a function of the STI alongside Hongisto's model (Jahncke *et al.*, 2013).

Jahncke, H., Hongisto, V. and Virjonen, P. (2013) 'Cognitive performance during irrelevant speech: Effects of speech intelligibility and office-task characteristics', *Applied Acoustics*, 74(3), pp. 307–316.

Figure 2.19 Subjective speech intelligibility as a function of the STI for sentences (non-optimized SRT) from which Hongisto's model (2005) was derived (IEC 60268-16, 2011).

IEC 60268-16 (2011) *Sound system equipment Part 16 : Objective rating of speech intelligibility by speech transmission index*. Brussels: European Committee for Electrotechnical Standardization.

Figure 2.20 Normalised preference levels of natural sounds used to mask two urban noises, road traffic noise, and construction noise (Jeon *et al.*, 2010).

Jeon, J. Y. *et al.* (2010) 'Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds.', *The Journal of the Acoustical Society of America*, 127(3), pp. 1357–66.

Figure 2.21 Visual images of water features and a street used by Jeon *et al.* (2012).

Jeon, J. Y. *et al.* (2012) 'Acoustical characteristics of water sounds for soundscape enhancement in urban open spaces.', *The Journal of the Acoustical Society of America*, 131(3), pp. 2101–9.

Figure 2.22 Spectra produced by different water flow rates. (a) Fountain (37 jets) with a 0.5 m extension. (b) Plain edge waterfall of 1 m width and 1 m height of falling water (Galbrun and Ali, 2013).

Galbrun, L. and Ali, T. T. (2013) 'Acoustical and perceptual assessment of water sounds and their use over road traffic noise.', *The Journal of the Acoustical Society of America*, 133(1), pp. 227–37.

Figure 2.23 Normalised preference levels of water sounds examined by Galbrun and Ali (2013).

Galbrun, L. and Ali, T. T. (2013) 'Acoustical and perceptual assessment of water sounds and their use over road traffic noise.', *The Journal of the Acoustical Society of America*, 133(1), pp. 227–37.

Figure 2.24 Semantic characterization of each water sound used over road traffic noise, illustrating both attributes and components (Galbrun and Calarco, 2014).

Galbrun, L. and Calarco, F. M. A. (2014) 'Audio-visual interaction and perceptual assessment of water features used over road traffic noise.', *The Journal of the Acoustical Society of America*, 136(5), pp. 2609–20.

Figure 2.25 Audio-only and audio-visual preference scores of water sounds combined with road traffic noise at an L_{Aeq} of 55 dBA (Jeon *et al.*, 2012).

Jeon, J. Y. *et al.* (2012) 'Acoustical characteristics of water sounds for soundscape enhancement in urban open spaces.', *The Journal of the Acoustical Society of America*, 131(3), pp. 2101–9.

Figure 2.26 Preferred water features from audio-only, visual-only, and audio-visual tests. (Galbrun and Ali, 2013).

Galbrun, L. and Ali, T. T. (2013) 'Acoustical and perceptual assessment of water sounds and their use over road traffic noise.', *The Journal of the Acoustical Society of America*, 133(1), pp. 227–37.

Figure 3.1 Laboratory structure used to generate the water sounds (Galbrun and Ali, 2013).

Galbrun, L. and Ali, T. T. (2013) 'Acoustical and perceptual assessment of water sounds and their use over road traffic noise.', *The Journal of the Acoustical Society of America*, 133(1), pp. 227–37.

Figure 3.2 Photographs of water features selected in this study, taken at the Building Services laboratory of Heriot-Watt University (Galbrun and Ali, 2013).

Galbrun, L. and Ali, T. T. (2013) 'Acoustical and perceptual assessment of water sounds and their use over road traffic noise.', *The Journal of the Acoustical Society of America*, 133(1), pp. 227–37.

Figure 5.3 Still images from animations of the water features used in the audio-visual condition of the preference and perception tests. Background image (Flickr: Space n.d.).

Space, K. (2013) *Arqiva_30-Open-plan office layout at Arqiva*, Flickr. Available at: <https://goo.gl/esDZcn> (Accessed: 15 March 2016).

Figure 5.13 Masking curves at 40 dB for two pure tones, 2.4 kHz and 3.5 kHz (Fletcher, 1953).

Fletcher, H. (1953) *Speech and Hearing in Communication*. Edited by J. B. Allen. New York: Acoustical Society of America.

Figure 5.14 Spectrogram of two plosives consonants, spoken between two vowel sounds.
(UCL, 2017).

UCL (2017) *Speech, hearing & phonetic sciences, Chapter 6, UCL Division of Psychology and Language Sciences*. Available at: <http://www.phon.ucl.ac.uk/courses/spsci/iss/week6.php> (Accessed: 23 August 2017).

Figure 5.15 Spectrogram of five fricative consonant, spoken between two vowel sounds.
(UCL, 2017).

UCL (2017) *Speech, hearing & phonetic sciences, Chapter 6, UCL Division of Psychology and Language Sciences*. Available at: <http://www.phon.ucl.ac.uk/courses/spsci/iss/week6.php> (Accessed: 23 August 2017).

LIST OF TABLES

Table 2.1 Intelligibility ratings as a function of the STI.	30
Table 2.2 Different water feature configurations examined by Galbrun and Ali (2013).	49
Table 2.3 Acoustic classifications and target values of open-plan offices.	60
Table 3.1 Acoustic and psychoacoustic parameters of the water sounds normalised at 55 dBA.	73
Table 5.1 Octave-band sound pressure levels of six water sounds as well as the speech signal used in the audio-only and audio-visual preference and perception tests.	101
Table 5.2 Acoustic characteristics of six water sounds as well as the speech signal used in the audio-only and audio-visual preference and perception tests.	101
Table 5.3 z-scores, <i>p</i> -values and effect sizes achieved by comparing audio-only and audio-visual perception scores using Wilcoxon test.	113
Table 5.4 Normalised audio-only and audio-visual scores in both preference and perception tests.	114
Table 5.5 Spearman's correlation (r_s) between preference and perception scores for each waterscape in both audio-only and audio-visual conditions.	115
Table 5.6 Chi-square (χ^2) test statistic and odds ratios between audio-only and audio- visual scores in the perception test.	118
Table 5.7 Audio-only and audio-visual preference scores and rankings of six waterscapes used in this study, from all participants retained for the analysis and from clusters obtained from hierarchical cluster analysis.	119
Table 5.8 Audio-only and audio-visual perception scores and rankings of six waterscapes used in this study, from all participants retained for the analysis and from clusters obtained from hierarchical cluster analysis.	122
Table 5.9 Spearman's correlation (r_s) between participants' preference scores in the audio-only condition and some acoustic and psychoacoustic measures. ...	125
Table 5.10 Spearman's correlation (r_s) between participants' perception scores in the audio-only condition and some acoustic and psychoacoustic measures.	125
Table 6.1 Cognitive tasks used in previous research in the context of performance and masking background noise.	141

Table 6.2 List of items/questions used in the questionnaire in the task performance experiment.....	150
Table 6.3 Order of tasks used in the task performance experiment	152
Table 6.4 Pearson's correlation coefficient (r) between selected variables measured in the task performance experiment.	184
Table 6.5 A summary of F -ratios, p -values, and effect sizes (r) obtained as a part of the statistical analysis of results in four cognitive tasks.	191
Table 7.1 Items added to Part 2 of the questionnaire concerning the water feature, and participants' responses to those items.	208
Table 7.2 Spearman's correlation (r_s) between participants' responses and their age, noise sensitivity, and background noise level at workstations.	211
Table 8.1 List of equipment used to measure the single-number quantities and their descriptions.	219
Table 8.2 The sound levels of normal effort speech at distance of 1 m in free field from the speaker and the A-weighting of octave bands.....	224
Table 8.3 Single-number quantities measured in Office 1 from two source positions.	233
Table 8.4 Single-number quantities measured in Office 2 from two source positions.	236

GLOSSARY OF SYMBOLS

A	A-weighting (dB)
	Equivalent absorption area (m^2)
A_0	Reference absorption area, 1 m^2
C	C-Weighting (dB)
C_1	Reference quantity correction (dB)
C_2	Radiation impedance correction (dB)
C_τ	Useful to Late Energy Ratio (dB)
$D_{2,S}$	Spatial attenuation of the A-weighted sound pressure level of speech (dB)
$D_{n,i}$	Attenuation at measurement point n , and octave band i (dB)
F	F -ratio
H	Kruskal-Wallis test statistic
Hz	Hertz
L_{10}	Noise level exceeded 10% of the time (dB)
L_{50}	Noise level exceeded 50% of the time (dB)
L_{90}	Noise level exceeded 90% of the time (dB)
L_A	A-weighted sound pressure level (dB)
$L_{A10} - L_{A90}$	Temporal variation in level (dB)
L_{Aeq}	A-weighted equivalent sound pressure level (dB)
$L_{AeqHigh}$	A-weighted equivalent sound pressure level of high frequencies (dB)
L_{AeqLow}	A-weighted equivalent sound pressure level of low frequencies (dB)
$L_{Aeq,T}$	A-weighted equivalent continuous noise level, over period T (dB)
L_{Amax}	A-weighted maximum sound pressure level (dB)
L_C	C-weighted sound pressure level (dB)
L_{Ceq}	C-weighted equivalent sound pressure level (dB)
$L_{Ceq} - L_{Aeq}$	Low frequency content (dB)
L_{eq}	Equivalent sound pressure level (dB)
L_{Fmax}	Maximum sound level over a fast time constant (dB)
L_{Fmin}	Minimum sound level over a fast time constant (dB)
L_n	Noise level exceeded $n\%$ of the time (dB)
L_p	Sound pressure level ($\text{dB } 2 \times 10^{-5} \text{ Pa}$)

$L_{p,A,S,n}$	A-weighted sound pressure level of speech at position n (dB)
$L_{p,B}$	Background noise level in octave bands (dB)
$L'_{pi(pre)}$	One-third-octave band time-averaged sound pressure level at the i th initial microphone position (dB)
$L_{pi(ST)}$	One-third-octave band time-averaged sound pressure level at the i th microphone position (dB)
$L_{p,Ls}$	Sound pressure level in octave bands (125 to 8000 Hz) of pink noise (dB)
$L_{p,Ls,n,i}$	Sound pressure level at the n selected microphone position, at octave band i (dB)
$L'_{pm(pre)}$	Arithmetical mean value of the one-third-octave band time-averaged sound pressure level at six initial microphone positions (dB)
$L_{p,S,4m}$	Sound pressure level of speech at 4 m from the speaker (dB)
$L_{p,S,1m}$	Octave-band sound pressure level at 1 m from the acoustic centre of the sound source in free field (dB)
$L_{p,S,n,i}$	Sound pressure level of normal speech at the measurement point n , and octave band i (dB)
$\overline{L_{p(ST)}}$	Mean time-averaged sound pressure level in the test room with the noise source under the test in operation (dB)
L_{Smax}	Maximum sound level over a slow time constant (dB)
L_{SNapp}	Apparent signal-to-noise ratio (dB)
$\overline{L_{SNapp}}$	Average apparent signal-to-noise ratio (dB)
L_W	Sound power level (dB re 10^{-12} Watts)
$L_{W,Ls,i}$	Sound power level of omnidirectional loudspeaker in each one-third octave band, under reference meteorological conditions
M	Mean
N	Sample size
	Total number of measurement positions
$N_{M(pre)}$	Initial number of microphone positions
N_M	Number of microphone positions
S	Surface area (m ²)
	Total surface area of reverberant test room (m ²)
S/N	Signal-to-noise ratio (dB)
$S/N (A)$	A-weighted Signal to Noise Ratio (dB)

T	Averaging time period (s)
T_{20}	Reverberation time measured for 20 dB drop, extrapolated to T_{60} (s)
T_{30}	Reverberation time measured for 30 dB drop, extrapolated to T_{60} (s)
T_{60}	Reverberation time (s)
U	Mann-Whitney test statistic
U_{τ}	Useful to Detrimental Energy Ratio (dB)
V	Pillai-Bartlett trace
V	Volume (m ³)
W	Watts
W	Kendall's coefficient of concordance
	Sound power (Watts)
W_0	Reference sound power level, 10 ⁻¹² W
c	Speed of sound (m/s)
f	Frequency (Hz)
	Resonant frequency of a bubble in an infinite volume of water (Hz)
f_m	Modulation frequency (Hz)
$h_k(t)$	Impulse response of octave band k
m	Modulation reduction
$m(f_m)$	Modulation reduction factor
p	p -value (probability)
p_0	Reference sound pressure level, 2×10^{-5} Pa
p_s	Static pressure (Pa)
$p_{s,0}$	Reference static pressure, 101325 Pa
r	Effect size
	Pearson's correlation coefficient
	Radius (m)
	Source-receiver distance (m)
	Water bubbles' radius (m)
r_0	Reference distance, 1 m
r_D	Distraction distance (m)
r_n	Distance to measurement position n (m)
r_P	Privacy distance (m)
r_s	Spearman's correlation coefficient
t	Time (s)

z	z -score
χ^2	Friedman's ANOVA test statistic Chi-square test statistic
θ	Air temperature, C°
θ_0	314 K
θ_1	296 K
AI	Articulation Index
AL_{CONS}	Articulation loss of consonants
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
BCa	Bias-Corrected and accelerated bootstrap
CI	Confidence intervals
CSH	Changing state hypothesis
ISE	Irrelevant speech effect
MANCOVA	Multivariate analysis of covariance
MANOVA	Multivariate analysis of variance
MTF	Modulation Transfer Function
Pa	Pascal, N/m ²
RA	Response accuracy (%)
RASTI	Room acoustic speech transmission index
RT	Reaction time (s)
SD	Standard deviation
SE	Standard error of mean
SIL	Sound interference level (dB)
SNR	Signal-to-noise ratio (dB)
SNR_k	Signal-to-noise ratio of octave band k (dB)
SPL	Sound pressure level (dB)
STI	Speech Transmission Index
df	Degrees of freedom

LIST OF CONFERENCES AND PUBLICATIONS

Abdalrahman, Z. and Galbrun, L. (2015) ‘Can Water Sounds Improve Work Spaces?’ poster presentation in *LoLo Conference*. Loughborough, United Kingdom, April 22nd.

Abdalrahman, Z. and Galbrun, L. (2017) ‘Soundscape assessment of a water feature used in an open-plan office’, in *PLEA: Design to thrive*. Edinburgh, United Kingdom, July 3-5.

Abdalrahman, Z. and Galbrun, L. (2017) ‘Audio-visual preferences of water features used in open-plan offices’, in *The 24th International Congress on Sound and Vibration (ICSV24)*. London, United Kingdom, July 23-27.

Abdalrahman, Z. and Galbrun, L. (2017) ‘Sound level and audio-visual preferences of water features used in open-plan offices’, to be submitted to *The Journal of the Acoustical Society of America* in May 2018.

Abdalrahman, Z. and Galbrun, L. (2017) ‘Cognitive performance effects and subjective satisfaction obtained from masking irrelevant speech with a water sound’, to be submitted to *Ergonomics* in June 2018.

Abdalrahman, Z. and Galbrun, L. (2017) ‘Assessment of how a water feature affects long-term subjective satisfaction and distraction distance in open-plan offices’, to be submitted to *Applied Acoustics* in August 2018.

CHAPTER 1: INTRODUCTION

1.1 General introduction

Open-plan offices are working spaces that make use of relatively large and open areas to accommodate a number of workstations and allow a group of people to work together. The idea of open working spaces was first addressed by Sir Stafford Northcote, and Sir Charles Trevelyan in a report presented to the UK Parliament in 1854, and later became the charter to the Civil Service (Northcote and Trevelyan, 1954). The report highlighted that for the “intellectual work” separate rooms are necessary while for the “more mechanical work” several people can work in the same room. The concept was later implemented in two projects in the United States, the Larkin Administration Building (1906) and the Johnson Wax company’s open-plan office (1939), both designed by Frank Lloyd Wright. The open-plan arrangement made its way back to Europe in the 1960s through the Bürolandschaft (Office landscape) movement in Germany. Since then, the concept has become popular in the design of workspaces across the globe.

An open-plan arrangement adds fluidity and flexibility to workspaces and allows for a greater integration of functions (Alfirevic and Simonovic, 2016). It accommodates a higher number of workers which will guarantee economic savings over private singular offices. It promotes communications and collaboration between co-workers which could be vital for certain types of works. However, not all types of work need this level of communication, interaction, and flexibility. As the Northcote-Trevelyan report suggested, intellectual works are better performed in private offices to avoid the interruption of the chain of thoughts of workers. Yet, economic advantages have made open-plan arrangements the norm in the design of modern work environments, irrespective of the type of work.

With the increasing popularity of open-plan offices, problems arising from this type of layout started to become more apparent. Early research on open-plan offices showed that workers internal motivation and satisfaction with work and colleagues decrease sharply after moving from private offices to an open-plan office (Oldham and Brass, 1979) and having little or no privacy can soon become a prominent issue in open-plan offices (Sundstrom *et al.*, 1982). Visual privacy can, to some extent, be provided through using opaque partition screens, while speech privacy is more difficult to obtain. The decrease

in speech privacy in open-plan offices arises from the lack of ability to carry out private conversations as well as being able to hear other colleagues' conversations. Lack of speech privacy in open-plan offices is a serious problem that has been shown not be offset by the favourable social climate that these spaces can provide (Hedge, 1982).

Returning to cellular private offices, despite providing privacy, might not be economically viable. Instead, there are acoustic solutions that can add a degree of speech privacy and thus attain some benefits of private office spaces without compromising on the qualities offered by open-plan offices. One solution, and arguably the most efficient, is using a masking system. Sound masking happens when a sound is played over another sound to render the latter less audible or inaudible. In open-plan offices, irrelevant speech is often considered the most annoying and distracting source of noise. Irrelevant speech is any speech that is not meant to be heard by the listener. Despite being irrelevant, the brain involuntarily processes information extracted from the speech which in turn has detrimental effects on subjective satisfaction and cognitive performance. Masking sounds have been shown to help reducing the detrimental effects associated with exposure to irrelevant speech in an open-plan office.

Many sounds in the literature have been used as speech masking sounds such as pseudorandom noises (i.e., white, pink and brown noises), babble and music. Due to their spectral and inherent positive features, recent studies have shown that water sounds can also be used in masking noise, especially road traffic noise (Galbrun and Ali, 2013; Galbrun and Calarco, 2014). The literature reports contradicting results on using water sounds as speech masking sounds in open-plan offices. Hence, it is difficult to draw conclusions regarding the effectiveness of water sounds in masking speech. This is mainly due to the lack of guidance and recommendations concerning the type of water sounds that should be used as speech masking sounds, that would be preferred and would positively affect cognitive performance and subjective satisfaction. This study, therefore, aims to fill this knowledge gap and contributes to the literature through providing recommendations specific to water sounds when used to mask irrelevant speech in open-plan offices. The recommendations include identifying the preferred masking sound level and spectra, as well as the effect of masking irrelevant speech with water sounds on cognitive performance and subjective satisfaction. Furthermore, the research examines the effect of visual stimuli on preferences of water sounds. The visual stimuli have not been examined in previous research in the context of open-plan offices, and hence, the need to fill this knowledge gap is evident.

1.2 Justification of the research

It is becoming increasingly difficult to ignore the constant noise issues arising in open-plan offices, particularly complaints about having little privacy. Despite the economic benefits, there is a growing body of scientific evidence showing that open-plan offices increase workers' dissatisfaction and cognitive workload (De Croon *et al.*, 2005), cause fatigue and difficulties in concentration (Pejtersen *et al.*, 2006) and result in subjective and objective impairment of work performance (Hongisto, 2005; Haapakangas, Haka, *et al.*, 2008; Haapakangas *et al.*, 2014). Dissatisfaction with the acoustic environment, i.e., background noise and lack of speech privacy, has repeatedly been highlighted as the main cause of the above problems (Sundstrom *et al.*, 1994; Jensen *et al.*, 2005; Bodin Danielsson and Bodin, 2009). Irrelevant speech coming from co-workers is the one particular factor that has been identified by numerous studies to have the most negative impact on the comfort level of workers, and having an appropriate speech masking sound, has been proven effective in mitigating the adverse effects of irrelevant speech (Hongisto, 2005, 2008; Venetjoki *et al.*, 2006; Virjonen *et al.*, 2007; Haapakangas, Helenius, *et al.*, 2008; Haapakangas *et al.*, 2011, 2017; Liebl *et al.*, 2016).

Several types of masking sound, mainly artificial sounds, have been used in the literature in the quest for finding the most efficient types of speech masking sound. Within that context, researchers have shown an increased interest in the use of water generated sounds for masking irrelevant speech, owing to the inherent positive and relaxing qualities of water sounds (Haapakangas *et al.*, 2011; Keus van de Poll *et al.*, 2015; Jahncke *et al.*, 2016; Hongisto *et al.*, 2017; Vassie and Richardson, 2017). The results, however, lack consistency when it comes to the effectiveness of water sounds in masking irrelevant speech and reducing the subjective workload. Some studies argue that water sounds are superior to artificially generated sounds (Haapakangas *et al.*, 2011; Keus van de Poll *et al.*, 2015), while Hongisto *et al.* (2017) found an artificially generated sound to be preferred over water sounds. Due to the discrepancy in results, further research is needed in view of identifying the preference levels of different water sounds and their effectiveness in masking speech. In addition, visual stimuli, which has been proven to play a key role in the perception of water features in urban soundscape studies (Watts *et al.*, 2009; Jeon *et al.*, 2012; Galbrun and Calarco 2014), have not been examined in the context of speech privacy in open-plan offices.

This research, therefore, aims at providing evidence-based solutions and guidance for using water features in open-plan offices in view of creating a pleasant work environment that provides speech privacy and promotes workers' cognitive performance and subjective satisfaction. Findings of the study will not be restricted to open-plan offices, as similar spaces, which suffer from noise could equally benefit from the study. Based on the above discussion, the need for conducting this research study is justified.

1.3 Research aim and objectives

The research is hypothesising that the use of highly rated water features in open-plan offices will be beneficial in creating a pleasant work environment that improves speech privacy. The main aim of the research is to provide evidence-based design solutions for water features used to promote acoustic satisfaction within open-plan offices and mask indoor noise, especially irrelevant speech coming from colleagues. To achieve this, the following objectives have been established:

Objective 1 - Identifying the preferred configurations of water features to be used in open-plan offices to promote speech privacy. This is achieved through the following sub-objectives:

- a) Identifying the preferred sound pressure level (SPL) of water sounds in open-plan offices as a function of speech intelligibility (STI) (Chapter 4).
- b) Identifying the preferred water features and their perception through audio-only and audio-visual tests (Chapter 5).

Objective 2 - Examining the extent to which water sounds improve cognitive task performance as well as subjective workload and satisfaction (Chapter 6).

Objective 3 - Examining the longer-term effects of water features on people's perceptions (Chapter 7).

Objective 4 - Evaluating the improvement in objective speech privacy associated with installing a water feature in open-plan offices through the measurement of distraction and privacy distances (Chapter 8).

Findings obtained will ultimately allow for providing evidence-based design solutions for open-plan offices in terms of improvements in speech privacy, acoustic satisfaction and cognitive performance, as well as reductions in workers' subjective workload.

1.4 Methodology

A variety of qualitative and quantitative experiments have been carried out to meet the aim and objectives of the study as thoroughly as possible. Initially, two experiments were carried out which included the identification of the preferred sound level of the water sounds, audio-only and audio-visual preferences of different water features and their effect on people's perception of their aural environment. These were followed by further experiments examining the effect of the preferred waterscapes on cognitive performance and subjective satisfaction, as well as the longer-term effect of placing a water feature in an open-plan office on worker's perception of their work environment. Human participation was vital to these tests, as it allowed for the objective measures to be subjectively validated. The last experiment was purely objective which involved measuring the distraction and privacy distances before and after adding a water sound to two open-plan offices. The water sound was selected based on the findings obtained in previous stages of the research.

1.4.1 *Experiment 1: Sound level preference test*

The first step towards meeting the objectives of this study was to determine the preferred sound pressure level of water sounds to be used in masking irrelevant speech in open-plan offices. Two water sounds were selected and played at five sound levels against a constant sound level of irrelevant speech. To account for the speech intelligibility, the irrelevant speech was played at two STI¹ levels. This allowed for the identification of the preferred masking sound level as a function of the speech intelligibility. Participants were invited to take part in this experiment where they compared all combinations of water sound levels using paired comparisons and stated their preferences. This experiment meets Objective 1 (a).

1.4.2 *Experiment 2: Audio-only and audio-visual preference and perception tests*

This experiment was designed to identify the type of waterscapes that would be preferred by people to mask irrelevant speech. The experiment also observed the change in people's perception associated with masking irrelevant speech with water sounds. Six water sounds were played at a constant sound pressure level determined in Experiment 1, against a constant sound level of irrelevant speech. Two test conditions were considered, an audio-only condition, and an audio-visual condition. The audio-only condition

¹ STI stands for the speech transmission index, which is a metric commonly used to objectively measure speech intelligibility. The metric is explained in Chapter 2.

included audio materials only, while the audio-visual condition included the water sounds and their visual representatives. Participants were invited to take part in the experiment where they compared the water sounds using paired comparisons and rated the change in their perception on a 5-point Likert scale. This experiment meets Objective 1 (b).

1.4.3 Experiment 3: Cognitive performance and subjective satisfaction

This experiment examined the effect of masking irrelevant speech with a water sound on cognitive performance and subjective satisfaction. Participants were invited to carry out four cognitive tasks under two background noise conditions, a speech-only condition and a masked speech condition. In the masked speech condition, irrelevant speech was masked with the preferred water sound played at the preferred sound pressure level, in line with the findings of Experiment 1 and Experiment 2. A satisfaction questionnaire was used to monitor the change in subjective satisfaction and workload after introducing the water sound. This experiment meets Objective 2.

1.4.4 Experiment 4 Longer-term exposure to a water sound

This experiment aimed at measuring the longer-term impact of a water feature on people's perception of their work environment. A water feature was purchased and modified to meet design criteria set out by previous experiments. This water feature was placed in a small-sized open-plan office for a period of three weeks. Satisfaction questionnaires were distributed before and after the installation of the water feature, to measure the rate of change in people's perception towards their physical work and sound environment. This experiment meets Objective 3.

1.4.5 Experiment 5 Objective measures

This experiment measured the rate of drop in the distraction distance, r_D , and the privacy distance, r_P , associated with installing a water feature in open-plan offices. Both r_D and r_P , alongside a number of other quantities, are known as the “single-number quantities²” recommended by BS EN ISO 3382-3 (2012) to represent room acoustic parameters in open-plan offices. The single-number quantities were measured in two open-plan offices before and after adding a water sound to the background noise, then the results were compared. This experiment provided an objective way to estimate what should be expected from installing a water feature in an open-plan office in terms of the reduction in r_D and r_P values.

² The single-number quantities are explained in Chapter 2 and Chapter 8.

1.5 Outline of thesis

Given the wide range of experiments conducted in this research project, the structure of the thesis mainly follows the experiments carried out. Each experiment is dedicated a chapter in which the methodology, the results, discussion and conclusions of that experiment are provided. This allows for fluidity of reading and makes navigation through the chapters easier. The review of literature is presented in one chapter, and final conclusions on the project are dedicated a chapter at the end of the thesis.

Chapter 2 provides a critical literature review in view of identifying the currently available knowledge and main findings, as well as any knowledge gaps that would need to be further investigated. The chapter identifies key studies that are of paramount importance to this research.

Chapter 3 briefly presents the various methodologies adopted throughout this research, alongside a detailed explanation of the statistical models that have been used in the data analyses. As mentioned earlier, this chapter only provides an overview of the methodologies. In-depth descriptions of the methodology of each experiment are given in their corresponding chapters.

Chapter 4 is dedicated to Experiment 1, in which the preferred sound pressure level of water sounds is identified, as well as the effect of the intelligibility level of irrelevant speech and the types of water sounds on the preferred sound level. Methodologies adopted in this experiment, analysis of data, discussion of the results and conclusions of Experiment 1 are all given in this chapter.

Chapter 5 is dedicated to Experiment 2, which is built on findings presented in Chapter 4. The chapter presents audio-only and audio-visual preferences of six water sounds, as well as their effects on people's perception when used as speech masking sounds. Methodologies and results of Experiments 2 are presented in this chapter followed by a critical discussion of the results. A comparison is drawn between the results presented in this chapter and results from relevant studies in view of providing a comprehensive understanding of the audio-only and audio-visual preferences of water sounds, in the context of speech masking in open-plan offices. Conclusions of the experiment are provided at the end of the chapter.

Chapter 6 is dedicated to Experiment 3, which presents methodologies, and results achieved in examining the effect of masking irrelevant speech on cognitive performance

and subjective satisfaction, when irrelevant speech is masked with a water sound. Advanced statistical analysis of data is performed in this chapter that includes examining effects of the gender of participants, nationality, and sensitivity to noise, as well as masking speech on cognitive performance and subjective workload. A critical discussion of results followed by the conclusions of findings are provided at the end of the chapter.

Chapter 7 is dedicated to Experiment 4, in which the longer-term effects of using a water feature in an open-plan office on people's perception and satisfaction level is presented. Descriptions of the water feature used in the experiment, the office space where the water feature was installed, and the questionnaires used to record people's responses are all explained in this chapter. Participants' responses to the questionnaire are presented before and after installing the water feature and comparisons are made to examine the magnitude of change in people's perception. Conclusions of the main findings of Experiment 4 are provided at the end of this chapter.

Chapter 8 is dedicated to Experiment 5, which aimed at measuring the rate of reduction of the distraction distance and privacy distance associated with adding a water sound to the background noise of two open-plan offices. The measurement procedures were performed in accordance with BS EN ISO 3382-3 (2012). Descriptions of the spaces where the test was carried out are provided in this chapter followed by results, discussion of results and conclusions of the main findings of Experiment 5.

Chapter 9 provides a summary of the conclusions, alongside the impact and implementation of the findings of the research in real life settings. The limitations of the study are also presented in this chapter, followed by recommendations for future studies.

Appendices A and B show the evaluation form and instructions used in Experiment 1, respectively. Appendices C and D show the evaluation form and instructions used in Experiment 2, respectively. Appendix E present general instruction given to participants in Experiment 3. Appendices F, G, H, and I illustrate the serial recall task, the one-back task, the information matching task, and the reading comprehension task, respectively. These tasks were part of Experiment 3. Appendix J consists of Part 1 of the satisfaction questionnaire used in Experiment 4 to measure peoples' responses before installing the water feature. Appendix K consists of Part 2 of the satisfaction questionnaire used in Experiment 4 to measure peoples' responses after installing the water feature. Finally, Appendix L tabulates people's responses to the items/questions included in the above questionnaires (i.e., Appendices J and K).

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

An extensive review of the currently available literature is presented in this chapter in view of providing key facts and findings, as well as identifying the knowledge gap which currently exists, on which this PhD research is founded. The chapter starts off by introducing fundamentals of acoustics which are essential for any soundscape study. An overview is provided on water features and their use in the built environment throughout history. The soundscape approach is then described alongside the advantages associated with using this approach to tackle noise issues in the built environment. A discussion is then provided on the inherent positive characteristics of water-related structures and their potential audio-only and audio-visual influences on the human perception.

The chapter also covers open-plan offices and their associated noise problems. The unconventional ways in dealing with noise complaints, through using masking sounds, is explained. Studies are reviewed in which water sounds have been effectively used as masking sounds for road traffic noise and irrelevant speech. Visually and acoustically preferred water features used in previous research are then reviewed in relation to their potential ability of providing speech privacy, tranquillity, peacefulness and relaxation, within the context of open-plan offices. The chapter also provides an explanation of speech intelligibility alongside the common objective metrics by which speech intelligibility is measured. The detrimental effect of irrelevant speech on cognitive performance and subjective workload are then highlighted. The chapter provides a critical discussion of the main findings of the relevant materials, resources and knowledge gap identified in the literature. Conclusions are drawn on the basis of these studies, in view of identifying suitable waterscapes that could be used as speech masking sounds in open-plan offices.

2.2 Psychoacoustics

The ear produces several different hearing sensations that cannot be explained using the physical acoustic measures such as the sound pressure level. Psychoacoustics is the scientific field of choice that bridges the gap between the physical measures and the hearing sensations (Fastl, 2006). Psychoacoustics is associated with the quality of sound,

it helps to identify whether a sound is perceived as being *rough*, *loud* or *sharp*. The psychoacoustic measures that are covered in this chapter are pitch and pitch strength, loudness, sharpness, roughness and fluctuation strength.

2.2.1 *Pitch and pitch strength*

Pitch is the sensation of how high or low a tone sounds, based on its relative position on a scale (Long 2014). The assigned unit of the pitch is *mel*, as the pitch is related to human sensation of melodies (Fastl and Zwicker, 2007). *Pitch strength* or tonality identifies how distinct a pure tone is in a complex noise (Kang, 2007). It describes how faint or strong a pitch is. It is possible for two sounds to share the same pitch but have different pitch strengths (Fastl and Zwicker, 2007).

2.2.2 *Loudness*

Loudness is the sensation of magnitude of a sound by the ear (Long, 2014). The ear does not respond similarly to different sound frequencies; the 2.7 cm auditory channel is partly responsible for perceiving the mid-high frequencies (2.7-3.4 kHz) as being louder (Long, 2014). There are two ways by which the loudness can be expressed, namely, loudness level, and relative loudness. The assigned unit of loudness level is *phon* which is the sound pressure level of a 1 kHz tone that is perceived as being as loud as the sound in question (Fastl and Zwicker, 2007). The relative loudness is measured using a relative comparison by asking subjects when a tone is twice/half as loud as another tone (Fastl and Zwicker, 2007). The assigned unit of relative loudness is *sone*, and the reference level is the 40-phon curve that is set to give a value of 1 sone. A sound having a relative loudness of 2 sones is twice as loud as a sound of 1 sone.

2.2.3 *Sharpness*

Sharpness is related to the high-frequency content in a sound. The higher the proportion of high frequencies, the “sharper” the sound becomes (Fastl and Zwicker, 2007). The key factors influencing sharpness are the spectral content and the centre frequency of a narrow-band sound. The assigned unit of sharpness is *acum*; a sound of 1 acum is defined as “a narrow-band noise one critical-band wide at a centre frequency of 1 kHz having a level of 60 dB” (Fastl and Zwicker, 2007, p. 239).

2.2.4 *Roughness and fluctuation strength*

Roughness is a complex effect that stands for the subjective perception of rapid (15-300 Hz) amplitude modulations of a sound. Amplitude modulations happen when a signal (low in frequency) makes a carrier wave (high in frequency) change its amplitude

according to the shape of the signal. The unit of measure is *asper*; one asper is the roughness of a 1 kHz tone played at 60 dB and 100% amplitude modulated at 70 Hz (Fastl and Zwicker, 2007). *Fluctuation strength*, on the other hand, is similar in concept to the roughness except it accounts for the subjective perception of a sound amplitude modulated at lower frequencies. The sensation of fluctuation strength remains up to 20 Hz, after which the sensation of roughness starts to take over. This transition is smooth and there is no clear border between the two sensations. Fluctuation strength is measured in *vacil*. A vacil corresponds to the fluctuation strength of a 1 kHz tone played at 60 dB and 100% modulated at 4 Hz (Fastl and Zwicker, 2007).

2.2.5 Weighting filters

The electrical weighting filters were developed to overcome the complexity of using loudness contour curves (Long, 2014). These filters are simply weighting factors, in dB, added to the original sound level to account for the frequencies at which the human ear is less sensitive (Peters *et al.*, 2011). Several filters can be found in the literature and they are designated by letters of the alphabet. The two most common weighting filters are the A-weighting and the C-weighting filters, which roughly correspond to mirrored copies of the 40 and 60 phone curves, respectively, as shown in Figure 2.1.

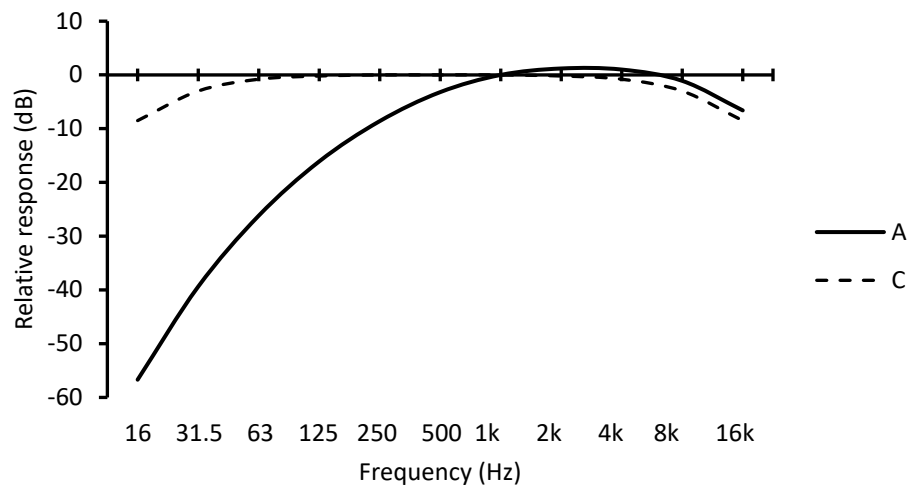


Figure 2.1 A-weighting and C-weighting curves (IEC 61672-1, 2013) (Fig. reproduced).

The A-weighting and the C-weighting filters can be applied to both the sound power level and the sound pressure level, and when they are applied, the unit of sound level becomes dBA and dBC for A-weighting and C-weighting filters, respectively. The value of $L_C - L_A$ is often used to estimate whether a broadband noise is dominated by low frequencies.

2.3 Exposure to natural environments

People simply prefer natural environments over built environments (Kaplan and Kaplan, 1989). Van den Berg *et al.* (2003) lay the reason for this to the inherent ability of the natural environments in providing restorativeness. They believe that the potential restorative capability of an environment makes it preferred over others. This belief was further supported by Staats *et al.* (2003), as they claim that natural environments have the ability to improve well-being and provide restoration from stress or attentional fatigue.

Exposure to a natural environment can have psychological benefits such as improved mood, well-being, and health (Hartig *et al.*, 2003; Laumann *et al.*, 2003). Some of the restorative effects associated with the exposure to natural environments are improved mood state, improved cognitive functioning and physiological signs of stress reduction (Ulrich, 1983; Ulrich *et al.*, 1991; Kaplan, 1995). Rooms overlooking vegetation or water, rather than buildings, have been reported to improve the attention capacity in adults (Tennessen and Cimprich, 1995) and the cognitive functioning in children (Wells, 2000).

Sound-wise, natural soundscapes can provide restorative benefits (Benfield *et al.*, 2014). Abbott (2015) found natural sounds to promote restoration and improve cognitive processes, while man-made sounds impeded restoration. Similarly, Jahncke *et al.* (2015) showed that natural sounds promote restoration, while office noises and broadband noises such as ventilation and traffic noises, do not.

There are two theoretical perspectives by which the natural environments' ability in providing psychological benefits can be explained. One approach is based on the capacity of natural environments to influence affectional states (Ulrich, 1983; Ulrich *et al.*, 1991), which assumes that the psychological benefits from exposure to a natural environment are directly associated with the stress-reducing capacity of that environment. The second approach regards the natural environment as being superior to the built environment, because of the former's higher attention restoration potentials. Everyday life requires a considerable amount of attentional recourses which may cause "attention fatigue". Nature has the capacity to offer involuntary attention which requires no effort, allowing the direct attention to rest (Kaplan and Kaplan, 1989; Kaplan, 1995). Involuntary attention is also known as "fascination" (Karmanov and Hamel, 2008). Kaplan (1995) identified four qualities that make the natural environment restorative. The qualities are *being away*, *fascination*, *extent* and *compatibility with human needs*. Kaplan (1995) argues that these values are not restricted to the natural environment and any urban environment that meets

the above qualities could be as restorative as the natural environment. This claim was empirically supported by Karmanov and Hamel (2008) when they examined the restorative capacity of natural and urban environments. Pre-stressed participants were shown video clips of a natural environment and an urban environment to investigate the restorative potentials of each environment. The study found that a well-designed and attractive urban environment can be as stress-reducing and mood-enhancing as an attractive natural environment. Interestingly, water was present in both video clips which could have played a key role in increasing the restorative capacity of those environments. However, the extent to which water increased this capacity remains unclear. Hence, adding elements of the natural environment (e.g., water) to the built environment could improve the restorative and relaxing potentials of the latter environment and benefit the occupants.

2.4 Water features

Streams, ponds, lakes, fountains, waterfalls, and cascades are all different types of water features. Examples of these water features can be found in nature as shown in Figure 2.2.



(a)



(b)



(c)



(d)

Figure 2.2 Examples of water features in nature. (a) Natural stream, Pagoeta Natural park, Spain. (b) Natural cascade, Mount Baranduda, Australia. (c) Niagara Falls, Canada. (d) Strokkur geyser in Iceland (Credit: Flickr).

The word *fountain* refers to a natural spring or a jet (or jets) or a spray of water. The very basic form of a manmade fountain consists of a jet and a basin. The jet pumps water into the air while the basin collects the falling water. Multiple jets and basins can be used to manipulate the water to the desired form. Several water configurations are possible as shown in Figure 2.3, by using different jet arrangements.

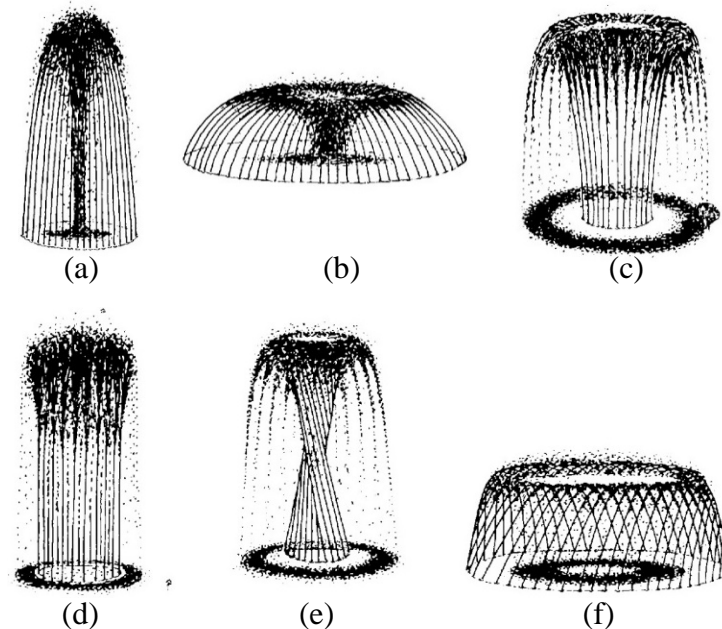


Figure 2.3 Different water patterns of fountains. (a) A dome with central drop. (b) A dome with inside drop. (c) A dome with outside drop. (d) Straight. (e) Basket weave outside. (f) Basket weave inside (Stephen, 1988).

A *waterfall* is simply water passing over a spillway (Stephen, 1988). It often incorporates features such as boulders, pools and vegetation (Brown and Rutherford, 1994). Various waterfall structures can be designed using different waterfall edges such as a plain edge, a saw-tooth edge and an edge made of small holes (e.g., Galbrun and Ali, 2013). In addition, the height of the waterfall, impact materials, the water flow rate and the collecting pool underneath can significantly influence the characteristics and the perception of waterfalls (Brown and Rutherford, 1994). Figure 2.4 shows an example of a waterfall in a built environment.



Figure 2.4 A waterfall in a built environment (Flickr: ~Pawsitive~Candie_N).

A *cascade* is water flowing over a series of steps with constant or variable vertical drops and horizontal extensions of the whole structure (Brown and Rutherford, 1994). It can be considered as a series of small waterfalls in which water flows from the highest waterfall down to the lowest. Similar to waterfalls, the height of steps, water flow rate, impact materials and the number of steps affect the design and perception of cascades (Brown and Rutherford, 1994). Figure 2.5 shows an example of a cascade in nature.



Figure 2.5 A multiple-step cascade in Mount Baranduda, Australia. (Flickr: Dirkus).

The water features explained above are mainly used in gardens and urban areas where there are no size restrictions. They are of a limited use for indoor spaces, due to their relatively large sizes. However, miniature versions of these water features with some modifications can overcome this problem and allow them to be used indoors. Two examples of water features that have been used in indoor spaces are shown in Figure 2.6.



(a)



(b)

Figure 2.6 Two examples of indoor water features. (a) Tokyo Disneyland Hotel Lobby's fountain (Flickr: Peter Lee). (b) A small pond-less indoor fountain (Flickr: Jay@MorphoLA).

2.4.1 Water sound generation

The process by which sound is generated as a result of water falling over water or other impact materials is complex. In the simple case of water drops falling onto a water surface, a low-level impact sound is generated followed by tonal sounds emitted by vibrating bubbles in the water. There is a short period after a droplet has contacted the water surface, during which, the contact region travels with a supersonic speed producing a small shockwave. The impact sound is believed to be caused by the supersonic shockwaves (Franz, 1959).

Bubbles are formed when air is trapped or injected into the water. The bubbles soon dissipate emitting a tonal sound which decays exponentially as energy is released back. Large bubbles can break up forming smaller bubbles which in turn produce several individual sound sources. The sound caused by the vibrating bubbles is believed to be the dominant sound. The emitted frequency depends on the resonant frequency of the bubbles which in turn is dependent on the bubbles' size (Leighton, 1994), as it is shown by Minnaert's formula (Minnaert, 1933):

$$f = \frac{3}{r} \quad (2.1)$$

where, f is the resonance frequency of the bubble and r is the radius of the bubble. Equation 2.4 shows that larger bubbles produce low frequency sounds while smaller bubbles produce high frequency sounds.

2.4.2 Water features and Architecture

Water features have been a key design element since the first gardens were tilled. Water features strongly influenced the shape of gardens as well as the pattern of life of the ancient Egyptians (Hopwood, 2009). Water features were also a prominent element in the Roman gardens.

Water features were an indispensable element in Islamic architecture. Water was extremely valuable in Muslim societies, and thus water features were usually placed at the very heart of gardens and courtyards, where attention could easily be drawn to (Lehrman, 1980). Fountains have helped Muslim designers to create a miniature version of the so-called “gardens of paradise” on earth (Carles *et al.*, 1992). Alhambra gardens in Spain, are prominent examples of the architect’s endeavour to create the gardens of paradise as shown in Figure 2.7.



Figure 2.7 Alhambra pool and fountains in Granada, Spain (Flickr: John Blower).

Similar to the Islamic architecture, Japanese and Chinese designers have made use of water features such as ponds, waterfalls, fountains and streams to create a “paradise on earth” type of environment (Dykstra, 2008).

The use of water in and around buildings has continued in the 20th and 21st centuries’ architecture. Fallingwater House (Figure 2.8), by Frank Lloyd Wright (1935-1937), is an example of the contemporary architecture in which water has been carefully blended with other rigid building elements.



Figure 2.8 Fallingwater House by Frank Lloyd Wright (Flickr: Nick Stanley).

The trend of using water in architecture will continue in future designs as water is essential to human life. “Some design elements and characteristics clearly span time and place. Order, space, form, texture, pattern, light and movement have no less relevance for today’s architect, artist and craftsman than they did many hundreds of years ago” (Lehrman, 1980, p. 7). A water feature is an element that has all these characteristics.

2.4.3 *The indispensable water*

Among the many elements that form the natural environment, water is probably the most prominent. Water can literally turn a desert into an oasis. Lehrman (1980) describes water as the element that offers a pleasant contrast to the rigidity and stability of the other architectural elements. Moving water is ever-changing and tireless, which brings life and delight to the environment. A few natural movements are as graceful and attractive as moving water (Burmil *et al.*, 1999). This movement is accompanied by sounds which make the experience even more exciting. The sounds made by water are constantly changing “no two raindrops sound alike, as the attentive ear will detect...of all sounds, the original life element has the most splendid symbolism” (Schafer, 1994, p. 170).

People are willing to pay more for houses and hotel rooms overlooking water. In a study by Luttik (2000), it was found that people were prepared to pay between 8% and 12% higher for houses with a view of water. Lange and Schaeffer (2001) reported similar figures for hotel rooms overlooking lakes as their rates were nearly 10% higher than rooms with views to a forest.

Out of 20 images of most tranquil urban environments, 17 images contained water in different forms (Pheasant *et al.*, 2008), and out of six posters used in an office to decrease

stress and anger at work, five were of aquatic scenes (Kweon *et al.*, 2008). Nordh *et al.* (2009) reported that small urban parks containing water features were perceived as being more restorative than ones with no water. Galindo and Corraliza (2000) reported that people tend to prefer those spaces that simultaneously contain *pleasure factors* (“comfort” and “pleasantness”) and *arousal factors* (“excitement/stimulation”, “being alive”); water has the capability to provide both factors (Faggi *et al.*, 2013).

Despite the growing body of evidence on water’s inherent positive effects; there are some studies in which water was found to have no effect in increasing restorativeness or preference levels. Ulrich *et al.* (1991) exposed pre-stressed people to video clips of natural environments with or without water. Physiological functioning such as blood pressure and skin conductance were measured in order to determine the ability of water scenes in helping people to get back to the normal stress level. The study found no statistically significant differences between the two settings. Similar results were reported by van den Berg *et al.* (2003) when they found natural environments to be superior to the built environment in providing restoration. However, the presence of water did not have any significant effect on perceived restoration, concentration, beauty and naturalness. In both studies, the videos that contained water were either dominated by vegetation and difficult to notice (Ulrich *et al.*, 1991), or the water was thin and dark brown in colour, which seemed unattractive (van den Berg *et al.*, 2003).

Water does not have a strong presence in modern indoor spaces, despite the scientific evidence that suggests the physiological and psychological benefits of water. White *et al.* (2010) argue that adding water to the built environment would significantly increase both preference and affective reaction ratings. Hence, bringing back this important element to modern indoor spaces would seem a step forward in the right direction, and this is what this PhD study has been aiming for.

2.5 A different approach to noise abatement

The conventional approaches to dealing with noise are mostly restricted to noise abatement strategies. Noise can be defined in various ways across different disciplines, however, the appropriate definition of noise for the purpose of this study is “unwanted sound”. This definition tends to be subjective and difficult to be effectively represented by quantitative measures such as the L_{Aeq} , which gives noise a meaning of “loud sounds”, and this is misleading, as not all noises are necessarily loud. For instance, in open-plan offices, the sound pressure level of normal speech is relatively low, yet regarded as an

annoying source of noise in many studies. Further explanations on noise in open-plan offices are given in the upcoming sections.

The ear has no “earlids” to protect it; the only mechanism by which the ear can protect itself is by filtering out undesirable sounds and concentrating on the desirable ones (Schafer, 1994). Hence, it would be rational to introduce desirable sounds into a soundscape where favourable sounds are scarce (e.g., in open-plan offices). This helps the ear concentrate on the favourite sounds and ignore the other undesirable ones. This is known as sound masking which occurs in two forms, namely, energetic masking, and information masking. The masking phenomenon is explained in Section 2.6. Therefore, to effectively approach noise problems in an environment where noise is disturbing, but not necessarily loud (e.g., open-plan offices), solutions should attempt at masking the noise rather than only reducing its sound pressure level. The masking phenomenon is discussed in detail in Section 2.6. Water generated sounds are generally regarded as desirable sounds and thus, could be used to help the ear protect itself from the sonic environment pollution.

2.5.1 *The soundscape approach*

Devices such as microphones, and sound level meters detect sound in a different way than how people do. Machines have no preference and they ignore the context in which a sound is played. They are designed to capture the acoustic environment which is defined by BS ISO 12913 (2014, p.1) as “sound at the receiver from all sound sources as modified by the environment”. Ultimately, the acoustic environment is perceived by people, not machines, and thus, the soundscape exists to address the human perception of the acoustic environment. The definition of soundscape is given in BS ISO 12913 (2014, p.1) as “acoustic environment as perceived or experienced and/or understood by a person or people, in context”. From the definition, it becomes clear that the soundscape involves dealing with human perception, interpretation of sound and the context, in addition to the acoustic environment.

The term soundscape is the auditory analogy of the visual landscape, as both are related to perceiving a physical phenomenon by people. It was first introduced in the 1970s by R. Murray Schafer, a Canadian composer (Schafer, 1994). The author contends that our soundscape has been polluted due to the overabundance of acoustic information and the pollution is not less important than the environmental pollution. Schafer emphasises that

a healthy soundscape cannot be created without discerning the sounds that enrich and feed listeners.

The soundscape is a process through which people can experience, understand and perceive their acoustic environment. This process, as highlighted in BS ISO 12913 (2014), consists of seven elements which are related to each other. The elements are: context, sound sources, acoustic environment, auditory sensation, interpretation of auditory sensation, responses, and outcomes, as illustrated in Figure 2.9.

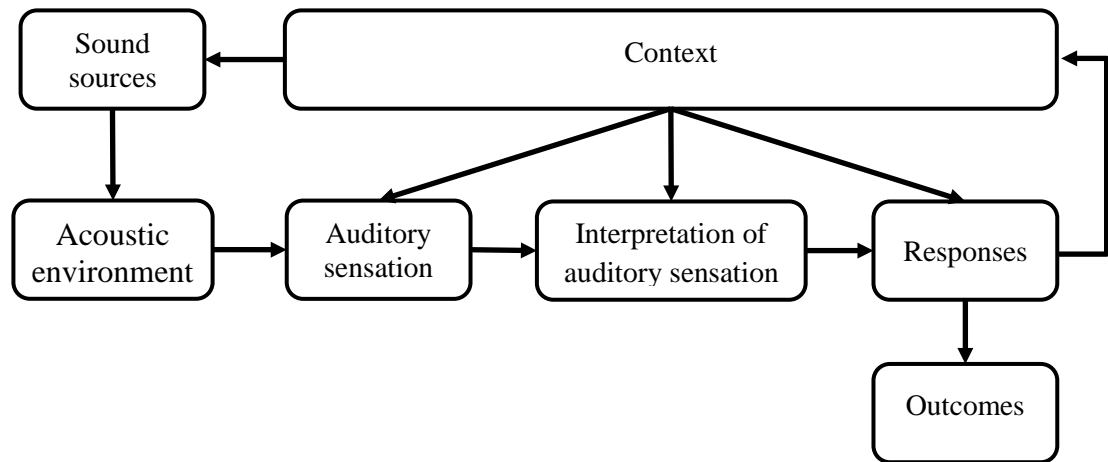


Figure 2.9 Elements in the perceptual construct of soundscape (BS ISO 12913, 2014).

As the figure shows, the context has a great influence on soundscape through affecting the auditory sensation, the interpretation of the auditory sensation and the response to the acoustic environment. On the other hand, responses, which include short-term reactions and emotions, influence the context. Finally, outcomes are long-term effects facilitated by the acoustic environment, such as attitudes, beliefs, judgments, habits, visitor/user experiences, health, well-being and quality of life, as well as reduced social costs for society (BS ISO 12913, 2014). Hence, soundscape studies are of a paramount importance in understanding how people evaluate and perceive their acoustic environment. As a result, the soundscape approach has been adopted in the current research, so that its findings are perceptually validated, which in turn allows for making more effective solutions, recommendations and suggestions to the noise issue that this study is addressing.

Soundscape studies on public spaces outnumber those dedicated to enclosed spaces. Understanding the soundscape environment in indoor spaces involves studying the sound propagation and attenuation in addition to the wave behaviour, i.e., reflection, refraction

and transmission should be addressed which are all acoustic characteristics of that space (Dokmeci and Kang, 2010).

The methods adopted in urban soundscape research should first be well understood, before studying the indoor soundscape. The urban soundscape study involves using methods and tools that are applicable to the indoor soundscape too. Dokmeci and Kang (2010) state that any indoor soundscape study should contain four main factors namely, objective factors, subjective factors, spatial factors and sonic factors, as shown in Figure 2.10.

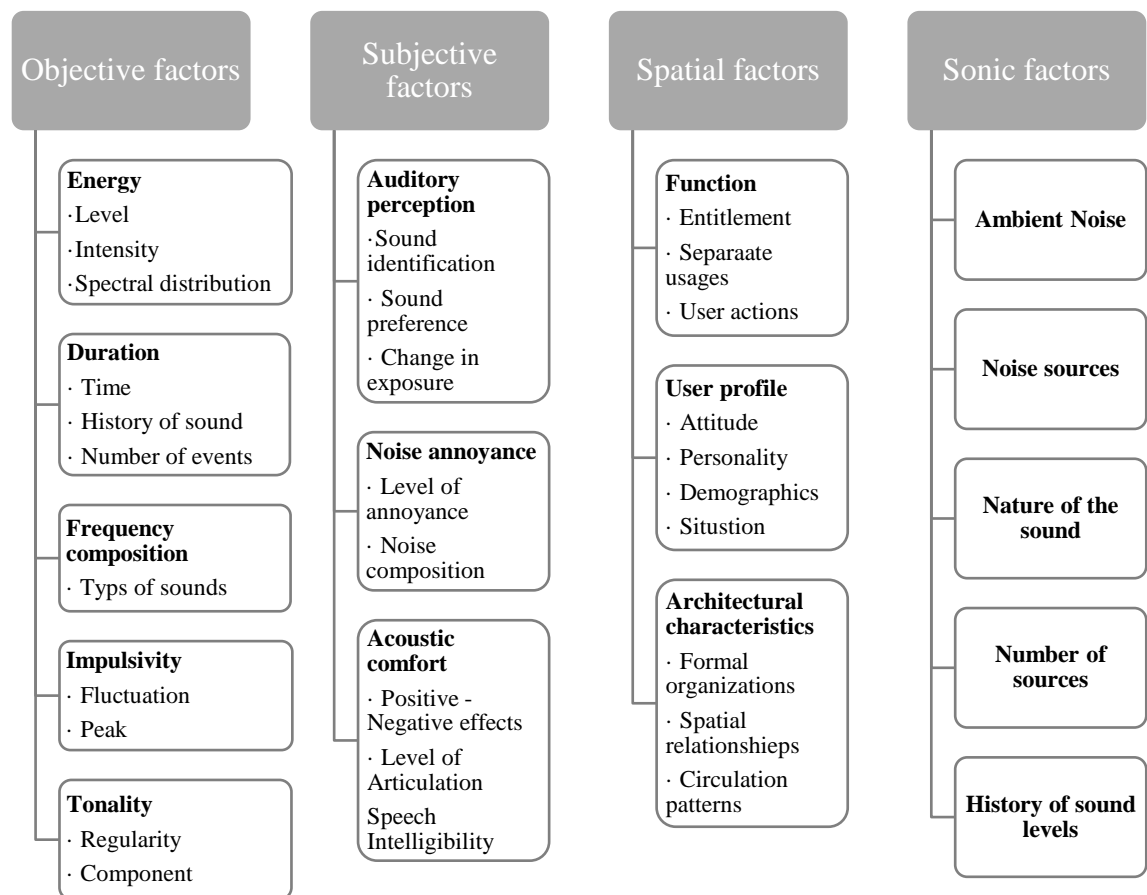


Figure 2.10 Factors to be considered in indoor soundscape studies (Dokmeci and Kang, 2010) (Fig. reproduced).

The *objective factors* depend mainly on the acoustic properties of the space in question as well as the sound source. Dokmeci and Kang (2010) recommend *objective factors* to be handled together with the *spatial factors* as they are all related to the type of space. Sound energy, duration, frequency, impulsivity and tonality are all part of the objective factors. The *subjective factors* involve dealing with the auditory perception, noise annoyance and acoustic comfort. These are usually examined using acoustic comfort

questionnaires or noise surveys. The *spatial factors* comprise the function of the building, user profile and architectural characteristics of the space. The *sonic environment* mainly involves dealing with the ambient noise in the space, individual sources, their nature and numbers as well as the history of the sound environment. Dokmeci and Kang's (2010) study provides a framework for soundscape research in indoor spaces, and thus the methodologies adopted throughout this research, are aiming at covering all four factors described above, to result in a comprehensive soundscape research.

2.5.2 Soundscape assessment

The ISO 15666 (2003) standard recommends using two types of questionnaires for the quantitative analysis of urban soundscapes in terms of annoyance. The two types of questionnaires are the 5-point verbal scale and the 11-point numerical scale. Jeon *et al.* (2010) examined the perceptual assessment of the quality of urban soundscapes. The study followed the ISO's recommendations (ISO 15666, 2003) by using both the 5-point verbal scale and the 11-point numerical scale. Broadly speaking, participants found the numerical scale easier for evaluating annoyance compared to the verbal scale. Both types of the scale have been used in this research, with more emphasis on the 11-point numerical scale. More justifications on this are given in Chapter 3.

2.6 Sound masking

Speech is highly disturbing in open-plan offices, and sound masking strategies have been proven effective in reducing the annoyance caused by speech. Sound masking happens when two tones having different sound levels are played simultaneously. The louder tone makes the quieter tone inaudible or less audible (Long, 2014). The auditory masking happens in two forms, *energetic masking* and *informational masking*. *Energetic masking occurs at the basilar membrane in the inner ear*, a masking sound (noise) makes a targeted sound (signal) inaudible or less audible by reducing the signal-to-noise ratio in the frequency regions of the targeted sound (Moore, 1995). *Informational masking* happens when an acoustic stimulus makes the brain exert more effort to process a target sound (Durlach *et al.*, 2003). This is the masking caused by the characteristics of the masker other than its energy. A typical example would be confusion due to the auditory similarity between the masker and the target sound such as a water sound (masker) and road traffic noise (target) (e.g., Axelsson *et al.*, 2014). If a part of the masking sound was perceived as the target sound, the masking effect of the masker would decrease. On the contrary, if a part of the targeted sound was confused with the masker, the masking effect

would increase. This phenomenon is known as the *target-masker confusion* (Durlach *et al.*, 2003; Watson, 2005). It has been shown that the decrease in the degree of similarity between the target sound and the masking sound significantly reduces the amount of informational masking affecting the target sound (Fastl and Zwicker, 2007).

Sound masking is technically increasing the background noise level which is associated with interference of sleep, annoyance, and the cardiovascular and psychophysiological systems (WHO, 2000). However, only a marginal increase in the A-weighted background noise level is expected when a masking sound is introduced. For instance, the level of noise in most open-plan offices is between 46 to 55 dBA (Navai and Veitch, 2003) and the recommended masking noise level is 45 dBA (Veitch *et al.*, 2002; Bradley, 2003). Adding such a masking sound would result in a marginal increase of the overall background noise level which is unlikely to have a detrimental effect on health.

2.6.1 Acoustic space

The acoustic space, defined by Schafer (1994), is the volume of space in which the sound of a sound source can be heard. The visual space and its corresponding acoustic space are congruous in a quiet environment where privacy is secured by effective barriers such as walls, fences and vegetation (e.g., private offices). However, in a noisy and less private environment, this congruence becomes less obvious and the acoustic space may no longer respect the visual space boundaries. For instance, the visual space dedicated to a person in an open-plan office is a workstation and its surrounding areas. The maximum acoustic space occupied by the person is the space in which his/her voice can be heard. Hence, the person's acoustic space will clearly exceed his/her visual space, causing interference with neighbouring acoustic spaces as shown in Figure 2.11 (a).

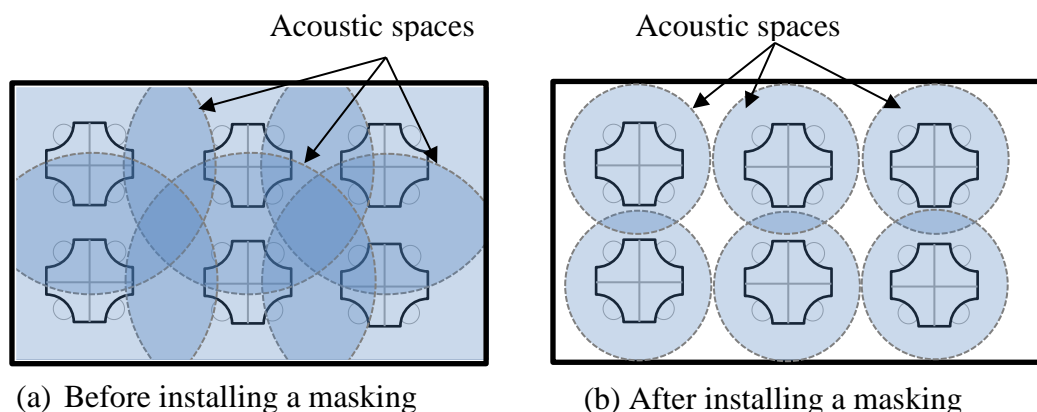


Figure 2.11 Acoustic spaces before (a) and after (b) installing a masking system.

Any masking sound, which is uninterrupted and not too loud, might become an acceptable *background noise* and might repress other intruding noises, making them being perceived to be quieter (Schafer, 1994). Some water generated sounds, air-conditioning noise, artificially generated masking noise (pink or white noise) are good examples of masking sounds.

Using water generated sounds as a masking system could be one possible solution for the problem of interfering acoustic spaces. The masking sound acts as a masker to noises especially irrelevant speech (see Section 2.9) coming from colleagues and makes them less audible or intelligible. This means that the acoustic spaces of people will become smaller and the congruence and consistency between the visual space and the acoustic space will become stronger, as shown in Figure 2.11 (b). In other words, people working in the space will be given the privacy of an enclosed office space without sacrificing the social and economic benefits of an open-plan office.

2.7 Human speech

Human speech is a disturbing source of noise in open-plan offices, which has detrimental effects on subjective satisfaction and cognitive performance (see Section 2.10). Therefore, understanding the acoustic properties of speech is important in devising solutions that would reduce its detrimental effects.

Human speech is a complex combination of sound waves, which cannot be represented via a simple sinusoidal signal. Speech patterns can be divided into two spectral components which are the audible spectrum and the modulation spectrum (Jacob, 2001). The first spectrum is produced by the vibration of the vocal cords during speaking (Jacob, 2001). This spectrum is characterised by containing more energy at the middle frequencies compared to the high and low frequencies. The modulation spectrum is generated when the audible spectrum passes through the articulators (teeth, tongue, lips and other speech organs). The modulation spectrum modulates the audible spectrum at various frequencies in order to generate phonetic information known as phonemes, from which words are formed (Long, 2014). The modulation spectrum can be divided into 14 frequency bands ranging from 0.63 Hz to 16 Hz at one-third octave intervals (Steeneken and Houtgast, 1980). Similar to the audible spectrum, the modulation spectrum is not flat and has its highest energy content at the middle frequencies. Figure 2.12 illustrates the audible (a) and the modulation (b) spectra of human speech.

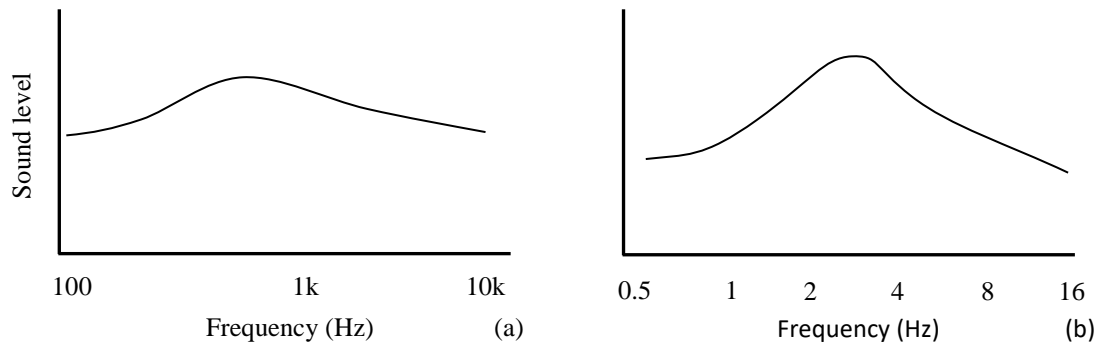


Figure 2.12 The audible spectrum (a), and the modulation spectrum (b) of human speech (Jacob, 2001) (Fig. reproduced).

The audible frequencies can be described as carrier waves on which the modulation frequencies or phonetic information are transmitted. This helps the phonetic information to travel through the air with less distortion, as the carrier waves are significantly higher in frequency, and therefore, less prone to be affected or distorted by the medium (Long, 2014). When the modulated carrier waves reach the listener, carriers and phonetic information is separated by the brain and the message will be delivered. The degree to which the phonetic information is distorted quantifies the level of understanding or the *speech intelligibility*.

2.8 Speech intelligibility

The degree of annoyance of speech is often related to its intelligibility level. Having a metric to measure the intelligibility level allows to quantify the annoyance caused by speech as well as the effectiveness of any masking system in reducing the intelligibility level of speech. The most accurate way for predicting speech intelligibility in a space is to record the subjective responses from listeners. Despite the accuracy of this method, it is considerably time-consuming and requires highly trained talkers and listeners. As a result, objective methods emerged.

The early attempts to improving intelligibility of telephone conversation were the basis for the metrics currently being used in predicting speech intelligibility in rooms. The common objective metrics are the *Articulation Index* (AI), the *Articulation Loss of Consonants* (AL_{CONS}), and the *Speech Transmission Index* (STI), as well as the *Useful to Detrimental Energy Ratio* (U_{τ}), the *Useful to Late Energy Ratio* (C_{τ}) and the *A-weighted Signal to Noise Ratio* (S/N (A)). Most of these speech intelligibility metrics are based on

the notion of “signal-to-noise” ratio, which is defined as the difference in dB, between a signal sound and the background noise. The extent to which a sound is considered as “signal” or as “noise” varies from a method to another. Among the objective metrics, the STI is widely accepted as a measure of the speech intelligibility due to its capability to account for speech distortion in both the frequency domain (i.e., interfering background noise) and the time domain (i.e., reverberation) (Steeneken and Houtgast, 1980). Besides, the STI is the recommended objective measure by BS EN ISO 3382-3 (2012) to represent acoustic properties of open-plan offices. Therefore, the STI has been adopted throughout this thesis as an objective measure of speech intelligibility.

2.8.1 Modulation transfer function

Houtgast and Steeneken (1973) first developed a system known as the Modulation Transfer Function (MTF), which could mimic many characteristics of human speech, and then introduced algorithms and worksheets needed to relate the MTF to speech intelligibility. They later proposed an objective speech intelligibility metric based on the concept of the MTF, termed the *Speech Transmission Index*, *STI* (Steeneken and Houtgast, 1980).

The STI requires an artificial speech-like signal which mimics the behaviour of human speech. To achieve this, the audible spectrum of speech can be replaced by a signal having a similar frequency spectrum to that of speech. This signal is then amplitude modulated at each standard octave-wide band by low-frequency tones known as the modulation frequency, f_m (Steeneken and Houtgast, 1980). The result is a sound wave with a spectrum similar to the graph shown in the left-hand side of Figure 2.13. Mathematically, it is expressed as a carrier wave being amplitude modulated by a sinusoidal function with a peak-to-peak amplitude of one (Houtgast and Steeneken, 1973).

When this amplitude modulated signal is transmitted through the air, similar to human speech, it faces distortion due to the background and reverberant noises. Background noise contributes to the distortion by raising the bottom of the signal above zero. Reverberant noise generates a delayed and probably deformed copy of the signal. A typical transmitted signal is illustrated in the right-hand side of Figure 2.13 which is clearly less modulated (i.e., more distorted) in comparison to the original signal, where the depth of the modulation envelope defines the degree of modulation.

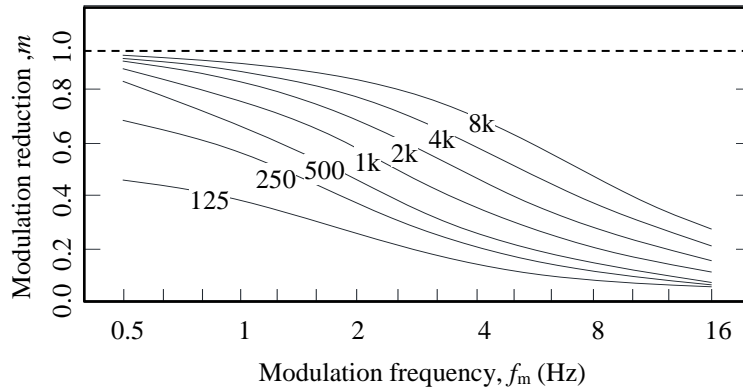
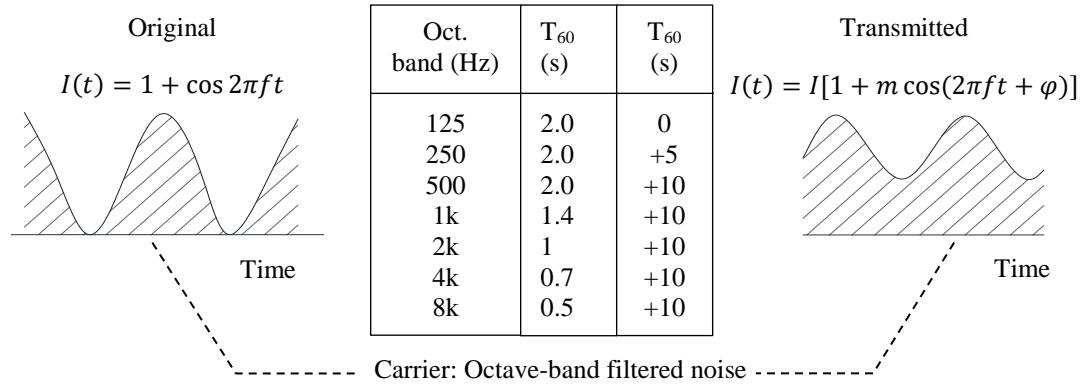


Figure 2.13 The Modulation Transfer Function (MTF) (Houtgast and Steeneken, 1985) (Fig. reproduced).

The modulation reduction, m , is described by the modulation reduction factor, $m(f_m)$, which is a function of the modulation frequency, f_m . The value of the modulation reduction factor varies between 0 and 1 for 100% and no reduction, respectively. The overall $m(f_m)$ can be mathematically calculated for an unamplified signal using the following equation (Houtgast and Steeneken, 1985)

$$m(f_m) = \frac{1}{\sqrt{1 + \left[2 \pi f_m \frac{T_{60}}{13.8}\right]^2}} \times \frac{1}{1 + 10^{(-0.1 L_{SN})}} \quad (2.2)$$

where $m(f_m)$ is the modulation reduction factor, L_{SN} is the signal-to-noise ratio (dB), f_m is modulation frequency (Hz), and T_{60} is the room reverberation time (s).

The first part of Equation 2.5 represents distortion due to reverberation, in which the modulation reduction has the form of a low pass filter with higher modulation frequencies being more affected by the reverberation as shown by the lower graph in Figure 2.13. The second part of Equation 2.5 stands for distortion due to the existence of background noise. Unlike distortion caused by reverberation, the background noise distortion is independent

of the modulation frequency, since the noise raises all levels within the carrier band, and thus, reduces the modulation equally (i.e., raises the bottom level above zero as explained before). The modulation reduction factor is measured using 14 one-third-octave bands starting from 0.63 Hz to 16 Hz, transmitted through 7 carrier signals each with an octave-wide frequency band from 125 Hz to 8 kHz, resulting in 98 (14×7) separate values of m .

Alternatively, $m(f_m)$, can be calculated from the impulse response of the space using Equation 2.6 (IEC 60268-16, 2011):

$$m(f_m) = \frac{|\int_0^\infty h_k(t) e^{-j2\pi f_m t} dt|}{\int_0^\infty h_k(t)^2 dt} \times [1 + 10^{-SNR_k/10}]^{-1} \quad (2.3)$$

where

- $h_k(t)$ is impulse response of octave band k ;
- f_m is the modulation frequency;
- SNR_k is the signal-to-noise ratio of octave band k , in dB.

2.8.2 Speech Transmission Index (STI)

The speech transmission index (STI) is a metric which makes use of the MTF to measure speech intelligibility. It was developed by Houtgast and Steeneken (1980, 1985) who provided a calculation scheme that transforms a set of m values into a single-number, STI, by using an apparent signal-to-noise ratio (L_{SNapp}) conveyed as a level. L_{SNapp} is defined as the signal-to-noise ratio that would have produced the modulation reduction factor, if all the distortion had been caused by noise interference, regardless of the actual cause of the distortion (Houtgast and Steeneken, 1985). L_{SNapp} is calculated using the following equation (Houtgast and Steeneken, 1985)

$$L_{SNapp} = 10 \log \frac{m}{1 - m} \quad (2.4)$$

where L_{SNapp} is the apparent signal-to-noise ratio (dB), and m is the modulation reduction factor. Limitations apply to the L_{SNapp} value. Any value greater than +15 dB is limited to +15. Similarly, any value lower than -15 dB is restricted to -15 dB. The average of all 98 apparent signal-to-noise ratios yields the STI after applying some weighting factors at each different octave bands as follows

$$\overline{L_{SNapp}} = \sum_{i=1}^7 w_i (L_{SNapp})_i \quad (2.5)$$

and

$$STI = \frac{[\overline{L_{SNapp}} + 15]}{30} \quad (2.6)$$

where $\overline{L_{SNapp}}$ is the average apparent signal to noise ratio (dB), w_i is weighting factors for octave bands from 125 Hz to 8 kHz (0.13, 0.14, 0.11, 0.12, 0.19, 0.17, and 0.14).

The STI is a number between 0 (no intelligibility) and 1 (perfect intelligibility). Intelligibility ratings was proposed by BS EN ISO 9921 (2003), which assigned subjective measures to ranges of STI. These intelligibility ratings are shown in Table 2.1. Speech privacy is the opposite of speech intelligibility, and hence inversely related to the STI. Therefore, in open-plan offices, low levels of the STI is desirable as they would ensure a higher level of speech privacy.

Table 2.1 Intelligibility ratings as a function of the STI (BS EN ISO 9921, 2003).

STI	<0.30	0.30-0.45	0.45-0.60	0.60-0.75	>0.75
Intelligibility rating	Bad	Poor	Fair	Good	Excellent

2.8.3 Factors influencing speech intelligibility

The two key factors influencing speech intelligibility in a room are room acoustics (e.g., reverberation time) and the signal-to-noise ratio of speech and ambient noise (Bradley *et al.*, 1999b). Bradley *et al.* (1999b) suggested that a signal-to-noise ratio of 15 dB and slightly less than 1.0 s reverberation time would be required to obtain 100% (excellent) speech intelligibility. Only slightly reduced intelligibility is expected for a signal-to-noise ratio of +10 dB (Bradley *et al.*, 1999b). Nearly impossible speech communication would be expected for ambient background noise of +60 dB, corresponding to less than -5 dB signal-to-noise ratio (Bradley *et al.*, 1999b).

Reverberation time, on the other hand, is considered as another influencing factor of speech intelligibility. Theoretically, the lower the reverberation time is, the better the intelligibility becomes (Bradley, 1986a, 1986b). However, the early reverberant sounds have some beneficial effects. The reflected sounds are important for increasing the sound level of speech, especially when the speaker is not facing a particular listener (Bradley,

1986b). Reverberation shorter than 40 ms positively contributes to speech intelligibility through increasing the loudness of speech (Bistafa and Bradley, 2000).

2.9 Speech masking in open-plan offices

In this study, the term *open-plan office* is used to describe floor plans, which make use of relatively large and open spaces and minimise the use of enclosed private rooms. An open-plan arrangement accommodates more people and allows easier communication, as well as reduces the space taken up by occupants, which guarantees economic savings. The expected economic advantages of open-plan offices are evident; however, these advantages do not come cheap. Open-plan offices have negative impacts on people who are working inside them. There is scientific evidence showing that open-plan offices increase workers' dissatisfaction and cognitive workload (De Croon *et al.*, 2005), cause fatigue and difficulties in concentration (Pejtersen *et al.*, 2006), and cause subjective impairment of work performance (Hongisto, 2005; Haapakangas, Helenius, *et al.*, 2008; Haapakangas *et al.*, 2014). Dissatisfaction with the acoustic environment, i.e., background noise and lack of speech privacy, has repeatedly been addressed as the main cause of the above problems (Sundstrom *et al.*, 1994; Jensen *et al.*, 2005; Bodin Danielsson and Bodin, 2009). Lamb and Kwok (2016) found that background noise in open-plan offices can have a detrimental effect on the self-rated workload and objective performance. They suggest that perceived thermal comfort, lighting comfort and noise annoyance can collectively account for between 2.4% to 5.8% reduction in performance for some cognitive tasks. They also found that these factors can act as a mediator indirectly affecting performance through negatively affecting motivation, tiredness and distraction levels which would, in turn, affect performance.

Irrelevant speech coming from co-workers is the one particular factor that has been identified by numerous studies to have the most negative impact on the comfort level of workers. This finding has been repeatedly reported in many studies, thus it is safe to state that little improvement would be achieved in the acoustic environment of an open-plan office without a thorough understanding and a proper treatment of this type of distraction (Hongisto, 2005, 2008; Venetjoki *et al.*, 2006; Virjonen *et al.*, 2007; Haapakangas, Haka, *et al.*, 2008; Haapakangas *et al.*, 2011; Liebl *et al.*, 2016).

Dealing with speech privacy complaints in open-plan offices requires numerous factors to be simultaneously taken into consideration, such as the ceiling absorption, the wall absorption, partitions, furniture, the height of separating screens and the distance between

workstations (Virjonen *et al.*, 2007). These factors are all related to the acoustic properties of the space. Introducing a masking sound is another possible solution which has been reported to be beneficial. This approach attempts at increasing the background noise level which subsequently decreases the signal-to-noise ratio of speech and makes it less intelligible, i.e., provides more speech privacy (Bradley, 2003; Virjonen *et al.*, 2007, 2009). The importance of speech masking was addressed at the early stages of open-plan offices, and since then speech masking has become a possible solution when the initial background noise level is not already exceeding 40 dBA (Veitch *et al.*, 2002; Haapakangas *et al.*, 2011).

Virjonen *et al.* (2007) studied factors affecting speech privacy between neighbouring workstations in an open-plan office. They examined several combinations of different heights and sound absorptions of the ceiling and the partition screens. They noticed that good speech privacy i.e., an STI of 0.5 (or less) in the neighbouring workstation, could only be achieved when the room acoustics and the speech masking were both taken into account. With 1.68 m high screens and a highly absorbent ceiling, a RASTI¹ value below 0.50 was achieved only when a pink noise with an L_{Aeq} of 48 dB was played as a masking sound. Without the masking sound, the value of RASTI remained above 0.70 even when a very high screen of 2.10 m was used. The study suggests that the positive effects of partitions and absorption in providing speech privacy in open-plan offices tend to increase with increasing sound pressure level of the masking sound.

2.9.1 Types of speech-masking sounds

Several types of sound have been used in the literature as speech masking sounds in open-plan offices. Examples of masking sounds used in previous research are pink noise (Ellermeier and Hellbruck, 1998), white noise (Loewen and Suedfeld, 1992) and filtered pink noise (Venetjoki *et al.*, 2006; Haka *et al.*, 2009). Music is another type of sound which is believed to have a mood-lifting influence on workers that continuous noise (e.g., pink noise) is incapable of (Oldham *et al.*, 1995).

Babble and non-speaking sounds caused by working in an open-plan office can provide a favourable masking, especially in large and reverberant offices. Babble is believed to have a positive masking effect in masking irrelevant speech (BS EN ISO 3382-3, 2012; Keus van de Poll *et al.*, 2015), yet Yadav *et al.* (2017) found babble to be beneficial only

¹ RASTI stands for the room acoustic speech transmission index, which was a shorter version of the full STI, but later became obsolete in IEC 60268-16: 2011.

when the number of voices that make up the babble exceeds 4 voices. They found babble made up by up to 4 voices to be more distracting than a single voice talker.

Recent studies have shown an interest in using natural sounds, especially water sounds, as a mean of masking irrelevant speech in open-plan offices, due to their inherent positive attributes and spectral properties (Haapakangas *et al.*, 2011; Keus van de Poll *et al.*, 2015; Jahncke *et al.*, 2016; Hongisto *et al.*, 2017; Vassie and Richardson, 2017).

(Haapakangas *et al.*, 2011) examined the effects of five different masking sounds on workers' cognitive performance. The masking sounds were: a filtered pink noise, ventilation noise, instrumental music, vocal music (music containing lyrics) and a spring water sound. Cognitive tasks examined in the study included a serial recall task, a creative thinking task and a proofreading task. For the serial recall tasks, all masking sounds, except the spring water, resulted in error rates significantly higher than the error rate in a silent condition. When results under the masking conditions were compared to a speech-only condition, all masking sounds, except vocal music, resulted in lower error rate, however, only the spring water sound resulted in significantly fewer error rates. The results are presented in Figure 2.14.

A marginal effect of the sound condition was observed in the creative thinking task, while performance in the proofreading task was not influenced by the sound conditions, confirming Haka *et al.* (2009) and contradicting both Venetjoki *et al.* (2006), and Smith-Jackson and Klein (2009).

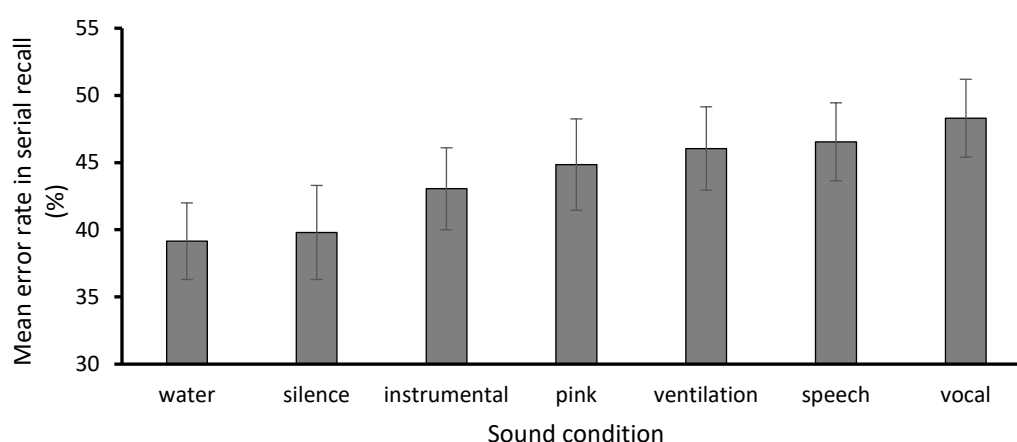


Figure 2.14 Mean error rates (%) in the serial recall task under seven sound conditions. Error bars represent the standard error (Haapakangas *et al.*, 2011) (Fig. reproduced).

Subjective responses revealed that speech and all masking sound conditions were significantly less satisfying than silence. In addition, participants showed more

satisfaction in all masking sound conditions compared to speech. These results confirm findings by Hongisto (2008) when he reported the self-rated work efficiency to increase in the presence of a masking sound. The study also supports the belief that irrelevant speech is often perceived as the most annoying source of noise. The satisfaction level for both the spring water sound and the filtered pink noise were similar and both were higher than that of the other three masking sounds, as shown in Figure 2.15. The spring water sound used in Haapakangas *et al.* (2011) was most beneficial in terms of both subjective (satisfaction) and objective (cognitive performance) indicators which none of the other masking sounds was capable of.

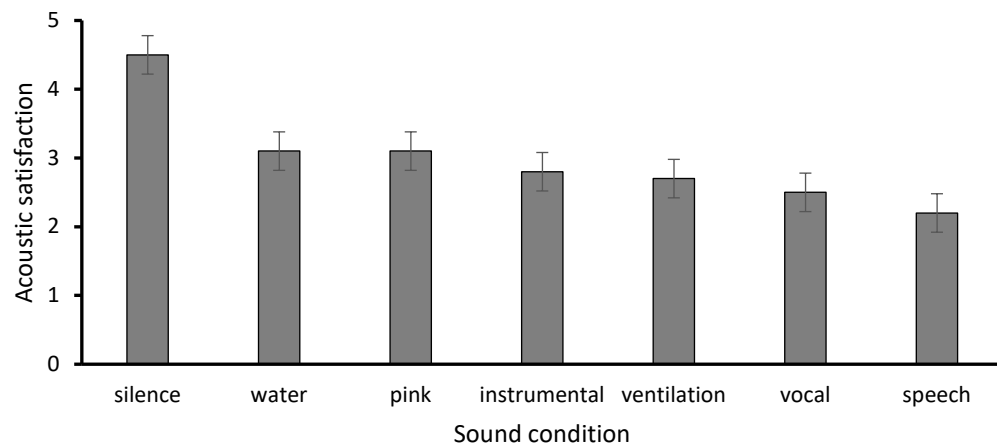


Figure 2.15 Acoustic satisfaction in the seven sound conditions. Error bars represent the standard errors (Haapakangas *et al.*, 2011) (Fig. reproduced).

Haapakangas *et al.* (2011) have no conclusive explanation about what made the spring water sound perform better than other masking sounds, but they lay the reason in the ability of water sounds to produce rapid level modulations which might have coincided with the fastest level modulation of speech at 8-15 Hz. However, the STI method used in this study only uses the equivalent sound pressure level of the background noise, ignoring the temporal variations of the masking sound. It is possible that the subjective speech privacy caused by the water sound could have been higher than that of the continuous masking noises such as the filtered pink noise and the ventilation noise, and thus, the higher satisfaction levels. Another possible explanation provided by Haapakangas *et al.* (2011) is the widely documented psychological benefits provided by natural environments to which the water sound belongs. The study suggests that more emphasis should be placed on the temporal characteristics of masking sounds instead of only measuring their equivalent sound pressure levels. A similar suggestion was made by

Galbrun and Ali (2013) in which they witnessed the temporal variation to be important in people's preferences towards water sounds.

A more recent study by Keus van de Poll *et al.* (2015) confirmed the above results in which they examined subjective satisfaction and cognitive performance (a serial recall task) of participants under five background noise conditions. The background noise conditions were silence, a single voice, a single voice masked by a pink noise, a single voice masked by water waves, and a single voice masked by multiple voices. The study found that the sound of water waves and the multiple voices (i.e., babble) significantly improved the performance level and the subjective satisfaction, in comparison to what was achieved under the single voice (i.e., speech-only) condition. Furthermore, both water waves' sound and multiple voices, proved to be more effective than the filtered pink noise in terms of improving the subjective satisfaction and the cognitive performance. The performance level when the water sound was used to mask speech was not significantly different than the performance in silent condition, confirming results achieved by Haapakangas *et al.* (2011). Multiple voices showed a similar result.

These findings were further supported by Jahncke *et al.* (2016), when they compared the cognitive performance and subjective workload in a serial recall task tested under five background noise conditions. The background noise conditions were quiet, speech-only, speech and headphones, speech masked by natural sounds, and speech masked by babble (7 voices). The natural sounds consisted of bird twitter and rippling water. The study found that cognitive performance under speech masked by natural sounds was not different than the cognitive performance in quiet, and was significantly better than the performance under the speech-only condition. Subjectively, while all masking sounds resulted in a significantly higher workload in comparison to quiet, natural sounds resulted in the lowest workload among the other masking sounds. The babble used in this study resulted in the lowest STI value (0.05), yet, was outperformed by the natural sounds which had an STI of 0.28. The study suggests that natural sounds are appropriate speech maskers, given that they restored performance level back to the baseline (i.e., performance in quiet) and shielded against workload.

The subjective responses of 18 employees under five masking conditions were compared in a recent study (Hongisto *et al.*, 2017). The masking conditions included four different water sounds and a pink noise. One of the water sounds was identical to the water sound used by Haapakangas *et al.* (2011). Unlike previous studies, the results showed that,

subjectively, none of the water sounds resulted in higher satisfaction levels than the pink noise. Therefore, the study could not recommend using water sounds over the pink noise as a mean of masking irrelevant speech in open-plan offices. The poor performance of water sounds in this study might be due to relying on energetic masking in choosing the water sounds. The spectra of the water sounds used in the study were close to the spectrum of human speech, as shown in Figure 2.16. However, previous studies (Watts *et al.*, 2009; Galbrun and Ali, 2013) suggest that the spectra of preferred water sounds do not necessarily match the spectrum of the masked noise.

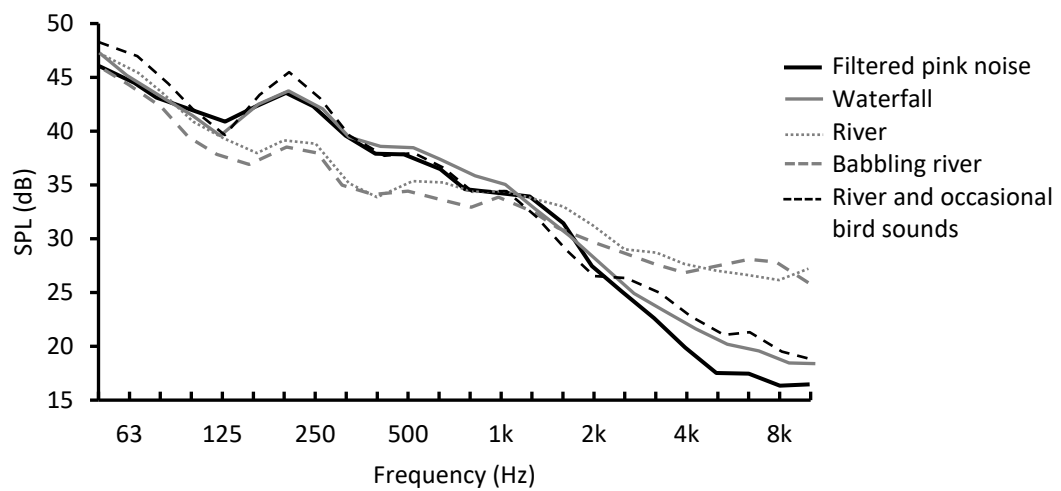


Figure 2.16 One-third octave band frequency spectra of four water sounds and a filtered pink noise used in Hongisto *et al.* (2017). (Fig. reproduced).

The spectra of the water sounds used in the study of Hongisto *et al.* (2017) are close to the spectra of sounds categorised as waterfalls by Galbrun and Ali (2013). Waterfalls are notorious for being disliked and poorly rated by people, mainly due to some semantic characteristics that make them being perceived as man-made sounds (Watts *et al.*, 2009; Galbrun and Calarco, 2014). This is further demonstrated by the qualitative descriptions given by participants in the study of Hongisto *et al.* (2017). The water sounds were described as “public toilet, fan and running water” which indicate how participants perceived the supposedly natural sounds as being man-made, and man-made sounds tend to be disliked (Watts *et al.*, 2009; Galbrun and Calarco, 2014). In addition, the water sounds were delivered via speakers that did not have a flat frequency response. Digital spectrum correction was performed to account for this shortcoming, which could have made the water sound lose their naturalness. This might also account for the discrepancy in results reported by Hongisto *et al.* (2017) and Haapakangas *et al.* (2011), despite using the same water sound.

2.9.2 Recommended masking spectrum and level

Veitch *et al.* (2002) suggest the optimum masking spectrum that keeps a balance between speech privacy and comfort level to be a pink noise whose sound pressure level reduces by 5 dB per octave band in the frequency range 125 Hz to 8 kHz. Hongisto *et al.* (2015) recommend a steeper masking spectrum, i.e., a spectrum of -7 to -9 dB per octave band increment, with respect to the general acoustic satisfaction. Both studies found masking sounds that are dominated by the middle and high frequencies to be the least satisfactory.

In terms of the masking sound level, Veitch *et al.* (2002) recommend the level of masking sound not to exceed 45 dBA. A similar recommendation was given by Haapakangas *et al.* (2011) when they suggested the A-weighted sound pressure level of a masking sound to be no more than 45 dB in order to allow for normal conversations between adjacent workstations and not less than 40 dB to produce an effective masking.

All recommendations and suggestions regarding the preferred masking spectra and levels are based on artificial masking sounds such as pink noise. No similar recommendations and guidelines could be found in the literature concerning water generated sounds. It appears that following these recommendations in choosing water sounds might not necessarily yield the expected result, as shown by the inferior performance of water sounds in Hongisto *et al.* (2017). Besides, the spectral properties are only one side of the picture; the evocative effects of water sounds have been shown to play a key role in people's preferences and perception towards water sounds (Watts *et al.*, 2009; Galbrun and Ali, 2013; Galbrun and Calarco, 2014). These effects seem to have been overlooked in previous research on masking sounds in open-plan offices (e.g., Haapakangas *et al.*, 2011; Hongisto *et al.*, 2017). Therefore, the need for more specific recommendations is evident in view of identifying appropriate water sounds that can be used in open-plan offices to mask irrelevant speech.

2.10 Effect of speech intelligibility on cognitive performance and satisfaction

The performance level of cognitive tasks is significantly affected in an environment where speech is highly intelligible. Typical examples of cognitive tasks may include mathematical, verbal, short-term memory and complex dual tasks. The performance is often determined by measuring the error rate. It has been reported that the error rate could increase between 4% and 41% in an environment dominated by speech compared to silence (Hongisto, 2005). The large variation is explained by differences in the

experimental design, such as speech type, the masking noise, the time pressure, the nature of the task, and the exposure time.

Irrelevant background speech can cause disruption in the short-term memory (Haka *et al.*, 2009; Jahncke *et al.*, 2013, 2016). The short-term memory is considered a crucial part of the human information processing system. The disruption of the short-term memory can reduce the cognitive performance by up to 30% (Beaman and Jones, 1997). It has also been found that irrelevant speech can impair more complex tasks such as proofreading (Venetjoki *et al.*, 2006; Smith-Jackson and Klein, 2009), recalling prior knowledge (Haka *et al.*, 2009) and logical reasoning (Landström *et al.*, 2002). On the other hand, there are studies which found no effect of the background noise on the performance in proofreading tasks, (Haka *et al.*, 2009; Haapakangas *et al.*, 2011) tasks requiring activation of the long-term memory (Jahncke *et al.*, 2013), reading comprehension (Haapakangas *et al.*, 2014), and the text memory (Haapakangas *et al.*, 2014). However, even in studies where no effect of irrelevant speech on cognitive performance was found, irrelevant speech had a detrimental effect on the subjective workload and satisfaction for most cognitive tasks.

Smith-Jackson and Klein (2009) investigated the effect of different types of background noise on proofreading tasks. They compared “quietness” (L_{Aeq} of 45-50 dB) to two irrelevant speech conditions, both having an L_{Aeq} of 65 dB. The two irrelevant speech conditions were intermittent speech and continuous speech. The study revealed that the completion rate and the false alarms were significantly affected by both speech conditions, in comparison to silence. In addition, the self-rated workload was significantly higher during the irrelevant speech conditions. The study suggests that even when the performance was not significantly affected, the subjective workload increased during the irrelevant speech conditions, which could eventually impair performance over longer periods of time.

Jahncke *et al.* (2013) examined five different cognitive tasks under various STI conditions ranging from 0.00 to 0.71. The cognitive tasks were a word memory task, an information search task, simple math tasks, a semantic word fluency task and a phonetic fluency task. The word memory task and the math task were significantly different between the baseline (i.e., STI 0.00) and STI 0.34. The performance decrement for each task was 5% and <3%, respectively, which confirm findings reported by Haka *et al.* (2009). No significant decrement in performance was recorded for the *Information search task* between the base line and STI 0.34. However, when STI 0.71 was considered, the

performance was significantly affected, and 6% reduction in performance level was recorded. This suggests that some tasks will not be affected by irrelevant speech until its intelligibility has exceeded a certain cut-off point. No significant effect of the acoustic condition was recorded for both the semantic and the phonetic word fluency tasks. This contradicts results reported by Haka *et al.* (2009) in which the long-term memory was significantly affected by irrelevant speech. The study suggests that participants could have used more effort to sustain the performance level, a claim which has been supported by Jahncke *et al.* (2011a) and Haapakangas *et al.* (2011) when they observed increased subjective workload during highly intelligible speech conditions. No conclusive evidence regarding participants' self-rating workload was provided to support this claim.

There is a discrepancy in findings reported by different studies on the effect of speech intelligibility on cognitive performance. This is largely due to the lack of a standardised way of measuring cognitive performance in open-plan offices, which has led to the emergence of numerous techniques by which the cognitive performance could be measured. Furthermore, it is believed that people would invest more effort to raise their performance level to the anticipated level when they are aware of a disturbance in their aural environment (e.g., increase in the subjective workload (Haapakangas *et al.*, 2011; Jahncke *et al.*, 2011a)). Ebissou *et al.* (2015) observed that participants whose performance was less affected by noise conditions reported a higher subjective workload, which suggests they might have exerted more effort to sustain a constant performance level. This would lead to the belief that even when reducing speech intelligibility does not result in a noticeable improvement in the performance over a short period of time, it might reduce the subjective workload and increase the satisfaction level.

The effect of the sound level of irrelevant speech on cognitive performance is believed to be marginal, in comparison to that of the intelligibility level. Schlittmeier *et al.* (2008) found subjective satisfaction and cognitive performance in a serial recall task to be significantly higher under low intelligible speech in comparison to that of high intelligible speech, despite both conditions having the same sound pressure levels. Performance in an arithmetic task under a low intelligible speech was not statistically different from that of silence, while a highly intelligible speech condition resulted in a performance level significantly lower than the level in silence. Subjectively, highly intelligible speech was perceived as being significantly more disturbing than low intelligible speech. No statistically significant differences were detected for a verbal logical reasoning, between low and high intelligibility conditions.

In another study by Jahncke *et al.* (2011a), the cognition, emotional and physiological effects of two noise conditions; high-level speech (L_{Aeq} of 51 dB) and low-level speech (L_{Aeq} of 39 dB) were investigated in simulated open-plan offices. Apart from different sound pressure levels, both noise conditions were similar in terms of the time variation, spectra, and the signal-to-noise ratio as well as the STI. A wide range of tasks was examined ranging from word and number recalling, response-ability, logical problem solving, math, serial recall, shifting and updating information and proofreading. Across all cognitive tasks, only word recalling was significantly affected by the high noise condition, while other tasks remained unaffected. Furthermore, no evidence was found to suggest that the stress hormones had increased during the high noise condition in comparison to the low noise condition. This finding suggests that the sound level of speech had little effect on the cognitive performance, in comparison to its intelligibility level (STI), confirming previous results reported by Hongisto (2005). Nevertheless, participants' self-ratings revealed that they were less motivated and more tired after 2 hours of work in the high noise condition. It is interesting that even with 12 dB reduction in the background noise, no significant increase in performance for most office tasks was detected, suggesting the need for a different approach in dealing with noise problems in open-plan offices, such as introducing masking sounds.

Using masking sounds to reduce the intelligibility level of speech has been shown to be an effective approach in reducing the detrimental effects of speech on cognitive performance and subjective workload. A growing body of scientific evidence suggests that cognitive performance and subjective responses of people working in open-plan offices can be improved through using appropriate masking systems that can reduce the intelligibility level of speech (Loewen and Suedfeld, 1992; Lewis *et al.*, 2003; Helenius and Hongisto, 2004; De Croon *et al.*, 2005; Haapakangas, Helenius, *et al.*, 2008; Smith-Jackson and Klein, 2009; Haapakangas *et al.*, 2011, 2014; Hongisto *et al.*, 2017).

In conclusion, despite some contradicting results, the cognitive performance in certain tasks, tends to be affected by irrelevant speech, especially the short-term memory tasks such as serial recall. However, the extent to which cognitive performance is affected is dependent on the type of task that is being tested and the experimental design adopted to measure the performance. In most studies where the effect of irrelevant speech on cognitive performance was absent, the subjective satisfaction and workload tended to be significantly affected by irrelevant speech. The results reported in the literature suggest that objective measures (i.e., error rate) and subjective measures (i.e., self-rated workload)

should both be used simultaneously to draw a more comprehensive picture of the effect of irrelevant speech on cognitive performance, within the context of open-plan offices.

2.10.1 Performance decrement as a function of the STI

It is well documented that irrelevant intelligible speech has a detrimental effect on cognitive performance for a range of cognitive tasks. This effect tends to increase when increasing the intelligibility level of irrelevant speech. Hongisto (2005) carried out an extensive literature review and proposed a mathematical model, shown in Figure 2.17, by which the decrement in task performance as a function of the STI can be estimated. The model explains the relationship between the STI and the performance decrement of some cognitive tasks such as serial recall, proofreading and reading comprehension tasks. According to the model, the performance decrement can increase by up to 7% in a condition where speech is perfectly intelligible. The model also shows that the cognitive performance remains unaffected until the STI has exceeded a value of 0.20, after which it starts to decrease rapidly. This trend continues until the STI has reached 0.50, where nearly the maximum performance impairment is expected to occur. Above an STI of 0.50, the speech intelligibility has a marginal effect on the performance decrement. Hence acoustic solutions in open-plan offices should aim at reducing the STI of irrelevant speech to a value below 0.50, if the performance decrement is to be reduced.

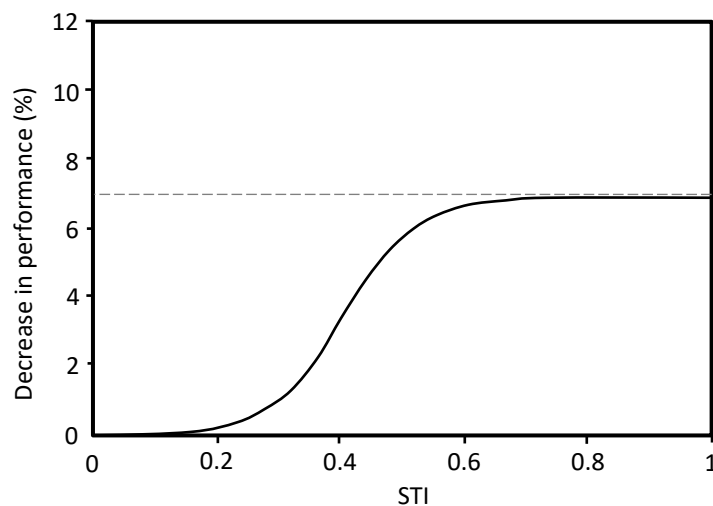


Figure 2.17 The schematic prediction model, which shows the decrease in performance as a function of the STI, irrespective of the sound level of speech (Hongisto, 2005) (Fig. reproduced).

The 7% decrement in performance in Hongisto's model (2005) was based on the minimum average decrement in task performance that Hongisto had found in the

literature, which could be safely generalized in such a way that all tasks are impaired by at least 7% by perfectly intelligible speech, in comparison to silence.

Venetjoki *et al.* (2006) examined the effect of various STI levels on work performance in laboratory conditions. The results supported Hongisto's model (2005), and found that the STI could be used as an indicator of the negative impact of irrelevant speech on the task performance. In another study by Jahncke *et al.* (2013), the accuracy of Hongisto's model (2005) was put under the question. They examined different cognitive tasks, under various STI conditions ranging from 0.00 to 0.71. The result confirmed that performance is affected differently under different STI conditions. The performance did not change in the STI range of 0.00 and 0.23, which is in agreement with Hongisto (2005). However, the maximum performance decrement was recorded at around STI 0.34, which disclaims the previous model, in which performance continued to be affected even after an STI of 0.50. Having said that, the general trend of the performance decrement curve, as shown in Figure 2.18, was similar to that of Hongisto (2005).

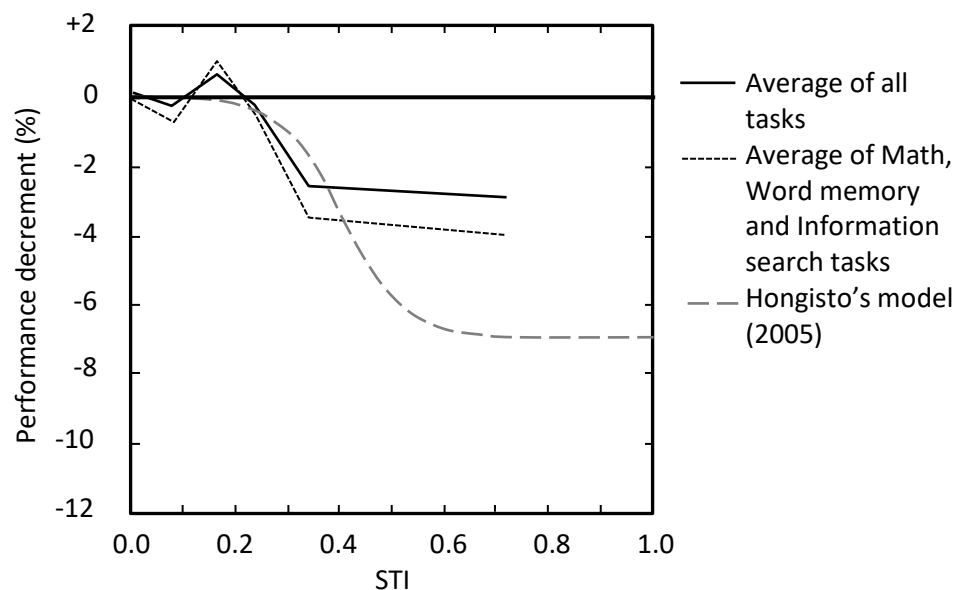


Figure 2.18 Task-averaged performance change between acoustic conditions as a function of STI alongside Hongisto's model (Jahncke *et al.*, 2013) (Fig. reproduced).

In a more recent study Ebissou *et al.* (2015), performance was mostly affected between STI 0.35 and 0.45, and the maximum decrement of performance was observed to have happened at an STI of 0.45. This plateau is somewhere in the middle between the STI 0.50 proposed by Hongisto (2005) and the STI 0.34 proposed by Jahncke *et al.* (2013). These differences in reported results by different studies are justifiable. The model

proposed by Hongisto was derived from the shape of the subjective speech intelligibility of sentences, as shown in Figure 2.19, and hence it represents the behaviour of performance decrement in relation to the STI, rather than the actual performance decrement levels. Therefore, higher or lower maximum percentages of performance decrement are expected for different types of task and noise conditions.

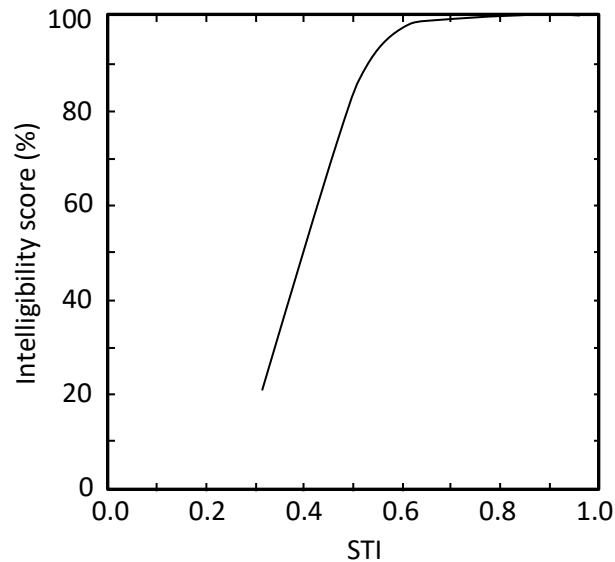


Figure 2.19 Subjective speech intelligibility as a function of STI for sentences (non-optimized SRT) from which Hongisto's model (2005) was derived (IEC 60268-16, 2011) (Fig. reproduced).

2.10.2 Theories behind performance decrement as a result of intelligible speech

There are two theories available in the literature explaining how irrelevant speech affects cognitive performance. These are the *Irrelevant Speech Effect* (ISE) by Salamé and Baddeley (1989) and the *Changing-State Hypothesis* (CSH) by Jones *et al.* (1992).

According to the ISE, it is the meaning of speech that disrupts memory and verbal tasks as they are similar, yet competing stimuli that need to be processed simultaneously. The CSH assumes that a decrement in performance occurs as a result of variations in speech which prevents the performers habituate to the acoustic environment. According to the latter theory (CSH), the meaning of speech does not have an impact on performance and thus, any irrelevant sound that has the same characteristic as varying speech would equally disrupt the cognitive processing. Both theories have their own supporters and opponents.

In support of the ISE, Smith-Jackson and Klein (2009) found that performance was more affected and workload was the highest during a continuous speech condition compared to

an intermittent speech condition. According to the CSH, the intermittent speech should have been more detrimental as it is more difficult to habituate to, due to its non-continuous nature. It has also been observed that the effect of speech on cognitive performance increases with increasing speech intelligibility (Hongisto, 2005). Furthermore, research has shown that dissatisfaction and self-reported workload increase with increasing speech intelligibility (Venetjoki *et al.*, 2006; Haapakangas *et al.*, 2017)

On the other hand, Banbury and Berry (1997) compared performance impairment levels under three different noise conditions, namely, continuous speech, speech repeated every three minutes, and random speech without a specific meaning. The performance was significantly affected by all noise conditions in comparison to silence; however, no significant differences were found among the three different noise scenarios which supports the CSH. Furthermore, Liebl *et al.* (2016) found continuous speech-like noise (no meaning) less detrimental on performance than variable-speech like noise (no meaning), which might partially lay support to the CSH.

Auditory distraction is believed to be caused when the information processing of different tasks in the brain conflicts with each other (Marsh *et al.*, 2008), such as when the involuntary processing of a sound (e.g., interpreting the meaning of speech) coincides with the intentional processing of a task, where both require the same information processing (e.g., interpreting the meaning of a written prose). Tasks, which require processing of order information (e.g., serial recall task), tend to be more vulnerable to unsteady noise conditions, since the brain needs to process the changes in the noise alongside the serial order processing involved in the tasks (Perham *et al.*, 2009). Furthermore, when semantic processing is required (e.g. reading a text, proofreading), performance tends to be more impaired by the meaning in speech, as it needs to be semantically processed too (Marsh *et al.*, 2008; Marsh and Jones, 2010). Morris and Jones (1990) proposed that disturbance caused by irrelevant speech is initially related to its meaning. However, when time passes, habituation occurs and further disruption would be caused by dishabituation (unexpected changes in the flow of speech). Scholars have not yet settled on the theory that is more accurate, however, there is a strong agreement among them that irrelevant speech does have a negative impact on cognitive performance which is a crucial point driving the current study.

2.11 Preference of water sounds

A limited number of studies used water sounds in open-plan offices, and these have been restricted to examining the effect of water sounds on cognitive performance and subjective satisfaction. No study has given recommendations and guidelines concerning the type and semantic characteristics of waterscapes that should be used within the context of an open-plan office. As a result, this section reviews studies in which water sounds and water features have been examined in urban spaces using the soundscape approach. Due to using the soundscape approach, these studies allow for a better understanding of the types of waterscape that tend to be preferred by people, regardless of the background noise that they are intended to mask.

Masking noise through the use of pleasant sounds has been considered as a potential solution for improving the acoustic environment in urban open spaces. Potential pleasant sounds are water sounds coming from water features erected in these spaces. Scholars have conducted many studies to investigate the effect of water sounds in urban open spaces within the context of the urban soundscape. Water sounds can play a complementary role to conventional noise mitigation strategies, due to the inherent positive and relaxing qualities (Kang, 2007) as well as their sound masking properties (Jeon *et al.*, 2010). Yang (2005) suggested that water in the form of fountains adds a colourful soundscape to urban open spaces, where road traffic noise is audible. Rådsten-Ekman *et al.* (2013) found the “pleasantness” of a sound environment dominated by road traffic noise to increase when highly pleasant water sounds such as sea sounds and stream sounds are added to the environment. Pleasant sounds were also found to increase the perceived “eventfulness” of the sound environment.

Watts *et al.* (2009) investigated the effectiveness of water generated sounds in masking road traffic noise and their potential ability in providing tranquillity. They found that there is a mismatch between the spectra of water sounds and road traffic noise. Generally, road traffic noise produces higher levels of low frequencies, which water sounds are incapable of masking. Watts *et al.* (2009) suggest that attempting to mask such low-frequency traffic noises with water sounds would require a much louder water sound that would, in turn, create a new noise problem. The study also found that water falling onto a cavity produces low-frequency sounds, while higher frequency sounds are associated with water falling onto hard surfaces, confirming previous work (Yang, 2005).

The most effective water sounds in improving tranquillity were sounds made by water falling over small boulders or combinations of boulders. Whilst, the least effective sounds were sounds made by water falling over variations of cavities. The results showed that the water sounds' ability in masking road traffic noises was not a decisive factor in how tranquil the water sounds were perceived. For example, water falling on a brick cavity received a very modest tranquillity rating despite producing the highest level of sound at the 250 Hz octave band, which was very close to the road traffic noise level at the same octave band. On the other hand, water falling onto a single small boulder produced a sound level at 250 Hz octave band which was 20 dB lower than the road traffic noise at the same frequency band, yet, was perceived as one of the top three most tranquil water sounds. In fact, the ability of water sounds in masking low frequencies was significantly and negatively correlated with the improvement in perceived tranquillity. Improvements in tranquillity were possible even when water sounds were played at levels around 7 dB lower than the road traffic noise level.

Tranquillity was also found to be positively related to the sound interference level (SIL), which is the arithmetical mean of the sound pressure level at the 500, 1k and 2k Hz octave bands. The study suggests that there is no point of producing water sounds with the correct spectral shape for masking low-frequency sounds while they deteriorate the perceived tranquillity in the space. The sound that improves tranquillity does not necessarily mask road traffic noise, but it is a pleasant sound which draws attention away from the more unpleasant traffic noise (i.e., provides informational masking rather than energetic masking).

The semantic tests revealed that *hollowness*, which was related to high levels of low frequency sounds, was perceived as a negative feature, while light temporal variations were considered a positive feature. In addition, the improvement in tranquillity was significantly correlated with water sounds that had been perceived as being *fast* and *varying*. Furthermore, sounds which evoked a sense of naturalness, e.g., *water falling as rain* and *water flowing over boulders in a stream*, were positively correlated with the improvement in tranquillity. Whilst, sounds perceived as being man-made, e.g., *water falling into a drain* and *water pouring into a container*, were negatively correlated with the improvement in tranquillity.

Among the psychoacoustic metrics, sharpness was found to be most closely correlated with changes in tranquillity. The sounds which contained more high-frequency contents

were perceived as being more tranquil. Loudness was less correlated with an inverse relationship with tranquillity. Roughness and fluctuation strength were not significantly correlated with the change in tranquillity.

The study suggests that to improve tranquillity while background noise is present, water features should generate natural sounds (e.g., water falling onto small boulders) which contain higher frequency variables and produce less low frequencies. The sounds should have relatively low-frequency contents and be variable in nature (temporal variations). A constant stream of water falling onto flat water or into cavities should be avoided as the sounds produced were associated with unnatural or man-made sounds.

Jeon *et al.* (2010) examined several natural sounds to overcome annoyance from two urban noises dominated by road traffic noise and construction noise. The natural sounds were a waterfall, rainfall, a stream, waves of a lake, birds in the forest, birds in a port, insects, the bell of a church, and wind. Sounds of the stream and waves of a lake were found to be the most preferred sounds to overcome both traffic and construction noises as shown in Figure 2.20.

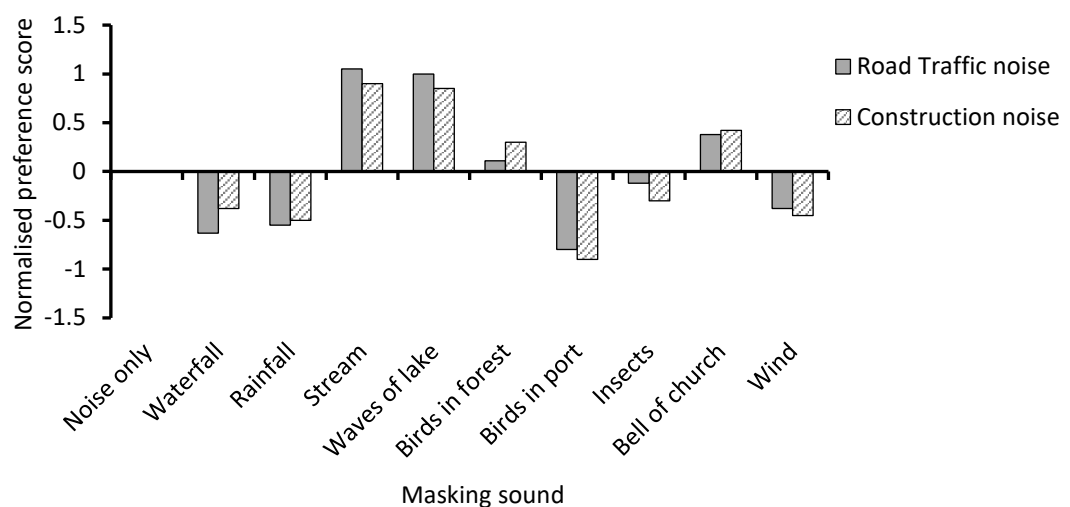


Figure 2.20 Normalised preference levels of natural sounds used to mask two urban noises, road traffic noise, and construction noise (Jeon *et al.*, 2010) (Fig. reproduced).

The figure shows that waterfall and rainfall sounds were both negatively rated, which is consistent with Watts *et al.* (2009) and would be confirmed further by Galbrun and Ali (2013). The study, therefore, suggests that the sounds of a stream and waves of a lake can be used as an effective natural sound to “mask” urban noises. The sound masking observed in the study, although not specified, was probably informational masking. Water

sounds, especially stream sounds, are not energetically effective maskers as they do not produce enough low frequency sounds to mask urban noises, which are dominated by low frequencies (Watts *et al.*, 2009; Galbrun and Ali, 2013). Besides, water sounds that produce high levels of lower frequency sounds (e.g., waterfalls) are often disliked. Therefore, as Watts *et al.* (2009) suggested, it is not the masking capability of water sounds that attracts people, it is their inherent positive evocative effects in providing peacefulness, tranquillity and naturalness. Having said that, introducing natural sounds would still reduce the loudness of road traffic noise due to the informational masking effects (Nilsson *et al.*, 2010). No spectra were given by Jeon *et al.* (2010) for the most and least preferred natural sounds, which would have given the opportunity to compare these findings with results from Watts *et al.* (2009), and Galbrun and Ali (2013).

Jeon *et al.* (2012) characterized the water sounds that could be used in urban open spaces for masking the road traffic noise. Sounds and visual images of 13 different water features were combined with road traffic noise. Participants were asked to rate each combination in terms of preference level. The results showed that all the combinations of water sounds and road traffic noise were significantly rated as being more preferred than road traffic noise alone. This supports the finding of a previous study by Jeon *et al.* (2010), in which it was stated that water sounds improve the perception of urban soundscapes. The visual images of water features used in the study are shown in Figure 2.21. The most preferred water sounds were fountain 2 (F2), stream 1 (S1), falling water 1 and 2 (FW1 and FW2). While the least preferred water sounds were fountain 3 (F3), fountain 4 (F4), stream 2 (S2) and waterfall 1 (W1). Part of these findings is in line with what Galbrun and Ali (2013) also found, in which stream and jet fountains were highly preferred. Conversely, waterfalls were very poorly rated in terms of preference and tranquillity in three studies (Watts *et al.*, 2009; Jeon *et al.*, 2010; Galbrun and Ali, 2013).

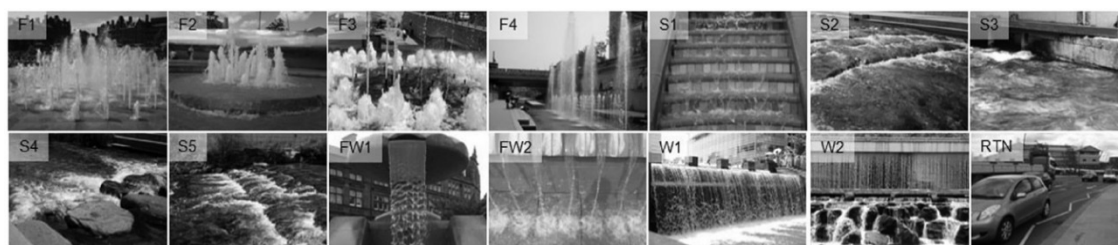


Figure 2.21 Visual images of water features and a street used by Jeon *et al.* (2012).

Jeon *et al.* (2012) also used a semantic scale for qualitative evaluations of the urban soundscape. Three semantic factors were identified, which were *freshness*, *calmness*, and

vibrancy. *Freshness* positively and significantly correlated with preference scores while *calmness* had a negative correlation with preferences scores, although the correlation was not statistically significant. This suggests that the fresher and the less calm the water sounds were perceived, the higher the preference scores were. The factor *vibrancy* had a weak ($r \leq .13$) and nonsignificant correlation with the preference scores.

Psychoacoustic analysis revealed that sharpness was positively correlated with preference scores. Furthermore, sharpness positively correlated with the factor *freshness* and negatively correlated with the factor *calmness*, indicating that higher sharpness is associated with water sounds being perceived as fresher and more energetic. The relationship between other psychoacoustic metrics and the preference levels was not statistically significant.

In another study (Hong and Jeon, 2013) on using natural sounds in masking road traffic noise, bird sounds and a stream sound significantly enhanced the preference level, whilst no marked improvement in preference was obtained when a waterfall sound was used to mask road traffic noise. This result is consistent with other research in which it was stated that natural sounds could improve urban soundscape and confirms previous results (Watts *et al.*, 2009; Jeon *et al.*, 2010; Galbrun and Ali, 2013) that waterfall sounds tend to be disliked, and therefore, should be avoided.

The keystone to the current study is a previous research by Galbrun and Ali (2013), in which numerous water features were examined using different design factors, such as the water flow rate, the height of falling water, waterfall's edge design and impact materials. Table 2.2 shows all the different water feature configurations tested in that study.

Table 2.2 Different water feature configurations examined by Galbrun and Ali (2013).

Code	Water feature type	Impact material	Category	Flow rate (l/min)
PEW	Plain edge waterfall	Water	Waterfall	120
SEW	Sawtooth edge waterfall	Water		30
SHW	Small holes waterfall	Water		30
SHC	Small holes waterfall	Concrete		30
FTW	Fountain (37 jets)	Water	Fountain	30
FTS	Fountain (37 jets)	Stone (pebbles)		30
DF	Dome fountain	Water		30
FF	Foam fountain	Stone and boulder		30
NJT	Narrow jet	Water		15
LJT	Large get	Water	Stream	15
CA	Cascade (4 steps)	Stone (pebbles)		15
ST	Stream	Stone and water		N/A

The water features were categorised under three groups; waterfall, fountain and stream. The study investigated how the acoustical and perceptual properties of different water features are influenced by their design factors. Different waterfalls, fountains, cascades and combination of upward jets were examined. Physical measures (spectrum and sound pressure levels) and psychoacoustic metrics (loudness, sharpness, roughness and pitch strength) were included in the measured data.

The frequency analysis of the water sounds showed the middle to high frequencies dominating the generated water sounds, with most of the sound energy confined to 500 Hz to 16 kHz octave bands, as shown in Figure 2.22 (a). Among the water features tested, the waterfall with a plain edge and a high flow rate produced the highest level of low frequency sounds as shown in Figure 2.22 (b).

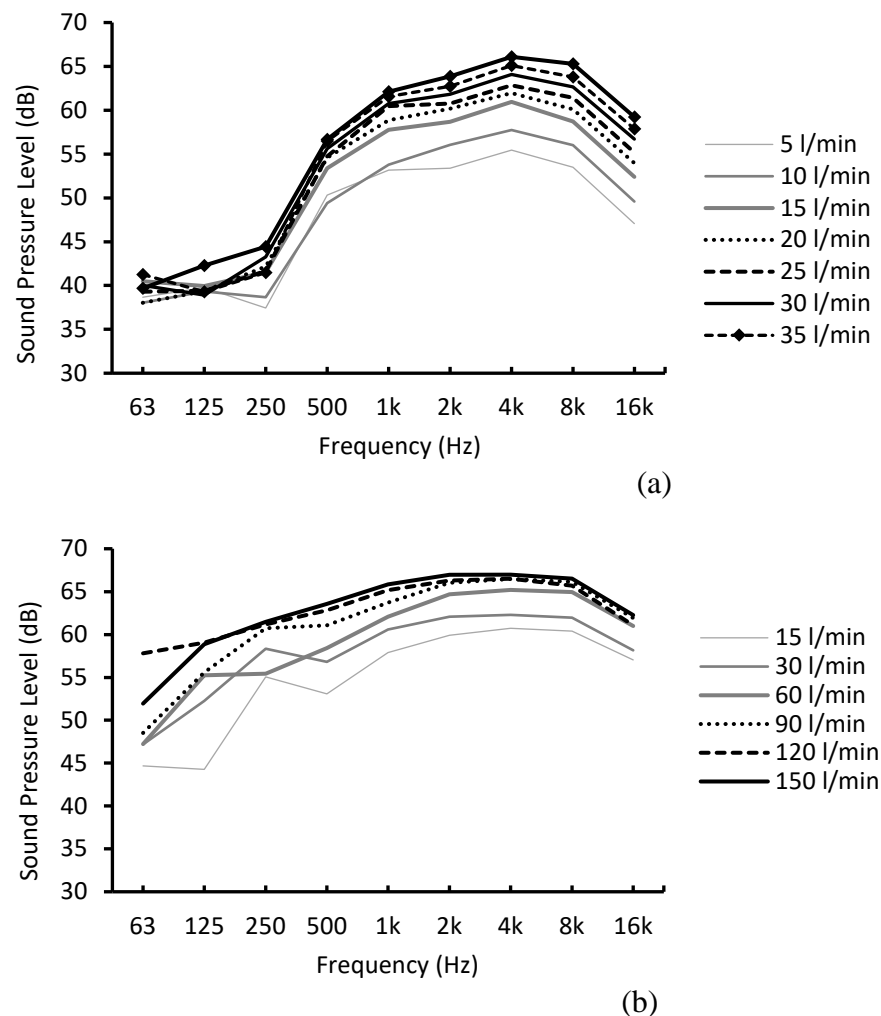


Figure 2.22 Spectra produced by different flow rates. (a) Fountain (37 jets) with a 0.5 m extension. (b) Plain edge waterfall of 1 m width and 1 m height of falling water. (Galbrun and Ali, 2013).

The laboratory test results of the study revealed that some acoustic properties of small to medium-sized water features can be anticipated based on the knowledge of their design factors. For instance, the equivalent continuous sound pressure level, L_{Aeq} , increased logarithmically with the flow rate for most types of water features. A similar trend was observed when loudness was examined instead of the L_{Aeq} . The changes in water flow rate equally affected all frequencies above 500 Hz, whilst lower frequencies were generally variable and less affected by changes in the water flow rate for all water features, except waterfalls. Furthermore, it was found that waterfalls could easily produce higher sound pressure levels in comparison to fountains, jets and cascades, and the produced sound had less temporal variations.

Impact materials were found to play a key role in determining the acoustic and psychoacoustic properties of the water sounds. The highest L_{Aeq} was recorded when water was used as an impact material, while plain solid surfaces such as concrete and metal resulted in lower L_{Aeq} (5-7 dB lower). Stone-like pebbles (30-60 mm), and gravel (10-20 mm) generated 2-4 dB higher L_{Aeq} than plain surfaces, but still lower than what was generated when water was used as an impact material. Boulders (150-250 mm) used over stones and gravel produced a very low L_{Aeq} (11 dB lower than water as an impact material). A combination of hard materials and water helped to increase the L_{Aeq} .

In terms of spectra, water as an impact material generated significantly higher sound pressure levels (typically 5 to 10 dB) at middle frequencies (250 Hz - 2 kHz), in comparison to the other impact materials. Water sounds generated by water falling over concrete, stones, boulders, and gravel were dominated by high-frequency contents. Differences among the impact materials were less prominent when the water flow rate was increased, and when the height of falling water increased.

Comparing the spectrum of road traffic noise to those obtained from the different water sounds revealed a mismatch between the spectrum of road traffic noise and the water sounds' spectra, which confirms results reported by Watts *et al.* (2009). Only a waterfall with large flow rates (150-200 l/min) could produce high levels of low frequencies comparable to that of road traffic noise.

In terms of the preferred water sounds, the natural stream (ST), the 37-jet fountain (FTW), the shallow jet with a low flow rate (LJT), and the four-step cascade (CA), were the most preferred water sounds. Whereas, the least preferred water sounds were the waterfalls

with small holes (SHW, and SHC), the waterfall with a plain edge and a large flow rate (PEW), and the single jet with a narrow nozzle (NJT), as shown in Figure 2.23.

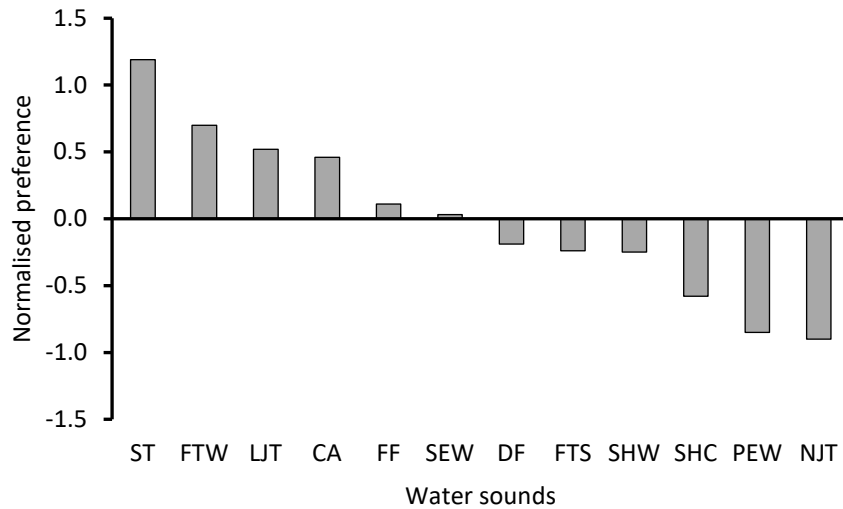


Figure 2.23 Normalised preference levels of water sounds examined by Galbrun and Ali (2013) (see Table 2.2 for definitions of acronyms) (Fig. reproduced).

In general, the statistical analysis showed that sounds categorized as *stream sounds* were preferred to *fountain sounds*, which in turn were preferred to *waterfall sounds*. This is interesting, since waterfalls whose spectra were most close to that of road traffic noise (i.e., had a good masking ability), were very poorly rated, confirming results reported by Watts *et al.* (2009), and suggesting that the energetic masking capability is not a decisive factor in identifying preferred water sounds.

The study also suggested that the preferred water sounds tended to have larger temporal variations ($L_{A10} - L_{A90}$), a larger low-frequency content ($L_{Ceq} - L_{Aeq}$), and a lower sharpness. No correlation with the roughness and pitch strength was found. These results are in line with results reported by Jeon *et al.* (2010) in which water sounds with a higher level of sharpness were associated with the factor “freshness” whilst lower sharpness was associated with the factor “calmness”. Water sounds with a high level of sharpness were preferred in Jeon *et al.* (2010) study, however, the perceptual analysis of the study was based on *freshness*, while in Galbrun and Ali’s (2013) study, the analysis was based on *peacefulness* and *relaxation*. In both studies, water sounds with low sharpness tended to promote relaxation. However, this finding contradicts the finding of Watts *et al.* (2009) in which water sounds with higher sharpness were more highly rated in terms of tranquillity. Galbrun and Ali (2013) tested a variety of upward and downward flows,

whilst Watts *et al.* (2009) examined only one downward stream with varying impact materials. If only results obtained from waterfalls were to be compared between the two studies, the contradiction disappears, as both studies found a positive correlation between sharpness and preference levels of waterfalls. Galbrun and Ali (2013) argue that this could be due to the fact that a downward stream with lower sharpness is often associated with man-made sounds such as water falling into a drain or container, and these tend to be disliked.

Water was found to be the preferred impact material. However, this was not necessarily true, as when different waterfall edges and impact materials were tested separately, boulders over stones were preferred to water. In addition, water was the least preferred impact material in the study of Watts *et al.* (2009). Therefore, the literature does not provide a clear-cut answer regarding the most preferred impact material. Yet, broadly speaking, a combination of water and boulders seems to be a good choice that would likely result in high levels of preference.

In another study by Galbrun and Calarco (2014), the audio-only and audio-visual effects of water features on human perception were investigated in the presence of road traffic noise. The audio materials used in the study were taken from Galbrun and Ali's study (2013). Audio-only preference tests revealed the preferred water sounds to be the natural stream (ST), the 37-jet fountains (FTW) and the four-step cascade (CA), whilst the least preferred water sounds were the small holes waterfalls (SHW), the narrow jet fountain (NJT) and the plain edge waterfalls (PEW) (refer to Table 2.2 for the list of water features). These findings are in line with Galbrun and Ali (2013), and somehow confirms results reported by Watts *et al.* (2009).

Preferred water sounds in the audio-visual tests were the natural stream (ST) followed by the four-step cascade (CA) and the 37-jet fountains (FTW), respectively. The least preferred water features were the large jet fountain (LJT), the plain edge waterfall (PEW) and the narrow jet (NJT). Broadly speaking, natural streams tended to be preferred to fountains which in turn were preferred to waterfalls, confirming Galbrun and Ali (2013). No significant correlations were found between preference scores and psychoacoustic metrics. Preference scores tended to be positively correlated with temporal variation, and negatively with sharpness, but with no statistically significant results.

Semantic analysis and categorization of the water sounds (audio-only condition) identified several attributes which were categorised under three key components, namely,

emotional assessment, *sound quality*, and *envelopment and temporal variation*, as shown in Figure 2.24. Statistical analysis showed that Component 1 significantly and positively correlated with preference scores, indicating that high value of *emotional assessment* was associated with high levels of preference. Component 2 showed a significant and negative correlation with preference scores. However, the negative correlation does not reflect poor sound quality. A negative correlation was observed simply because of high scores in perceived sharpness, perceived roughness, and speed, i.e., water sounds which had high levels of perceived sharpness, perceived roughness and speed, tended to be disliked.

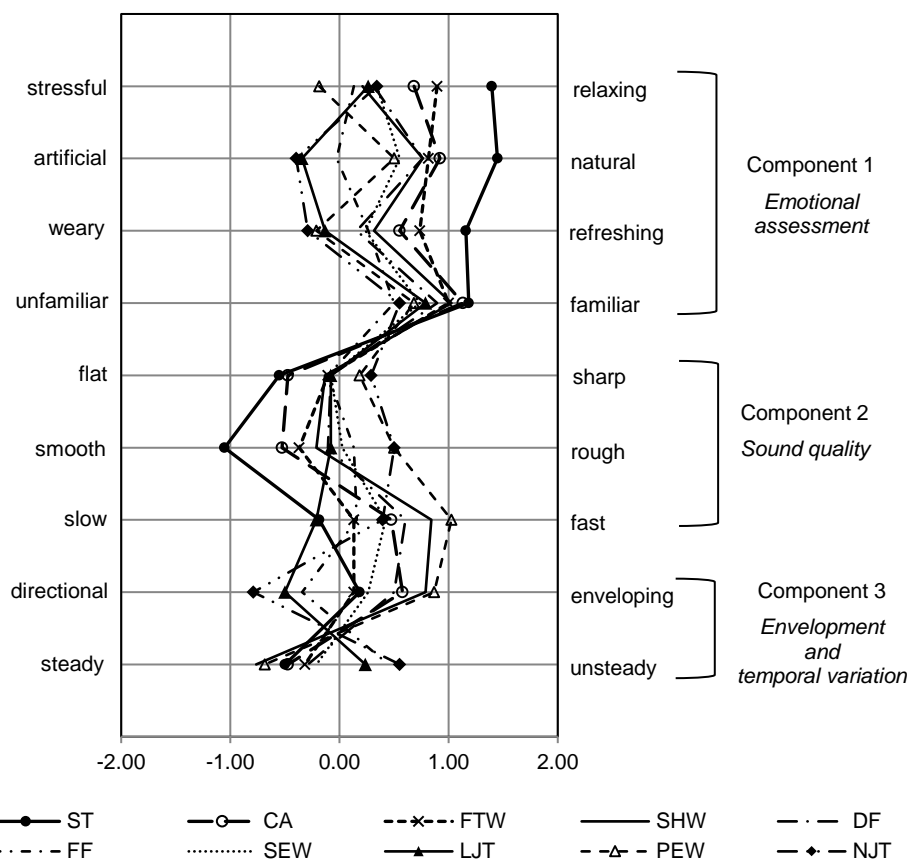


Figure 2.24 Semantic characterization of each water sound used over road traffic noise, illustrating both attributes and components. Results are given as average scores obtained for each attribute (refer to Table 2.2 for definitions of acronyms) (Galbrun and Calarco, 2014).

The study also suggested that the perception of waterscapes depends mainly on the emotional attributes associated with the sound (Component 1) and the characterization of

the sound quality (Component 2) whilst no significant impact of envelopment and temporal variation (Component 3) was recorded for the audio-only preferences.

Correlations between semantic components and acoustic/psychoacoustic parameters revealed that participants were unable to correctly identify the sharpness, roughness and temporal variations of the water sounds i.e., no correlation was found between physical parameters and their corresponding perceptual descriptors.

It is worth mentioning that pitch strength was significantly correlated with Component 1 while roughness was correlated with Component 2. This suggests that more attention should be paid to pitch strength and roughness in future waterscape studies. In addition, the psychoacoustic metrics tend to have a poor correlation with the direct preference or tranquillity rating (Watts *et al.*, 2009; Galbrun and Ali, 2013).

Evocation and qualitative categorization tests revealed that water sounds perceived as being natural were preferred to man-made sounds (e.g., water tap sounds). However, unlike Watts *et al.* (2009) the correlation was not significant. Similarly, correlation between the visual preference levels and the natural looking water features was positive while a negative correlation was observed between the visual preference and the man-made looking features. However, in neither conditions, the correlation was statistically significant.

An interesting finding of this study is that the natural stream sounds were reported as being easily identifiable, unlike waterfall and fountain sounds. Considering these findings and the *target masker confusion* theory (Durlach *et al.*, 2003; Watson, 2005) (see Chapter 2, Section 2.6), it could be concluded that water sounds which are easily identified as being natural, such as stream sounds, tend to improve the soundscape quality regardless of the background noise's type or level. Whereas, water sounds which are hard to be identified as natural sounds might be confused with the background noise, resulting in lower preference levels (e.g., Axelsson *et al.*, 2014). This could explain the discrepancy in preference scores in different studies regarding fountain and waterfall sounds; whilst stream sounds have consistently been reported to improve soundscape when road traffic noise is audible (Watts *et al.*, 2009; Jeon *et al.*, 2010; You *et al.*, 2010; Galbrun and Ali, 2013; Rådsten-Ekman *et al.*, 2013; Galbrun and Calarco, 2014).

Galbrun and Ali's study (2013) alongside research by Watts *et al.* (2009), Jeon *et al.* (2010), Jeon *et al.* (2012), and Galbrun and Calarco (2014) provide a deep level of understanding of the acoustic and psychoacoustic properties of different water features,

as well as people's perceptions of different water sounds. Despite these studies being conducted within the context of the urban soundscape, their results could still be implemented with some limitations in indoor spaces, as urban soundscape studies involve using methods and tools that are often applicable to the indoor soundscape. Thus, these studies offer a pool of data on which the current study is founded.

2.12 Audio-visual preferences of water features

The audio-visual preferences of waterscapes in urban spaces have been extensively investigated by numerous studies. However, very few studies have examined the audio-only preferences of natural sounds in open-plan offices, and only one study has been found in the literature on the audio-visual preferences of water generated sounds in open-plan offices. As a result of this lack of research in the area, this section mainly covers research investigating audio-visual preferences of water sounds in outdoor spaces such as parks, and plazas, which could give an understanding of the role that visual stimuli can play in changing people's perception and preference level toward a particular water sound.

Carles *et al.* (1992) used a combination of 4 different recorded sounds, namely, bird songs, water sounds, cricket and park noise, in addition to images of 8 different visual landscapes (32 combinations altogether) to investigate the audio-visual interaction and soundscape preferences of these settings. The audio-visual preferences were found to be more influenced by the audio stimuli. The study also found that when the visual stimuli and the aural stimuli are congruent, preference levels tend to increase. For instance, a combination containing the sound of water and an image of a tropical river was significantly more preferred than a combination of the sound of water and an image of a barren dry land. This suggests that when people hear a particular sound, they will likely expect to see the environment from which the sound is coming.

In their study on masking traffic noise with water sounds, Watts *et al.* (2009) found that experiments which contained a video clip of a water feature were considered as being more tranquil and preferred over other configurations in which the visual stimulus was removed (i.e., audio-only conditions). Jeon *et al.* (2012) observed visual stimuli to have a significant effect on preference levels of water sounds. They evaluated people's preference of road traffic noise masked by 13 combinations of water sounds, under two test conditions, an audio-only condition and an audio-visual condition. The addition of the visual stimuli resulted in an increase in preference scores for all water sounds except

falling water 1 and 2 (F1 and F2) as shown in Figure 2.25. The figure also shows that the water sounds benefited differently from the addition of the visual materials, a result which would be further demonstrated by Galbrun and Calarco (2014).

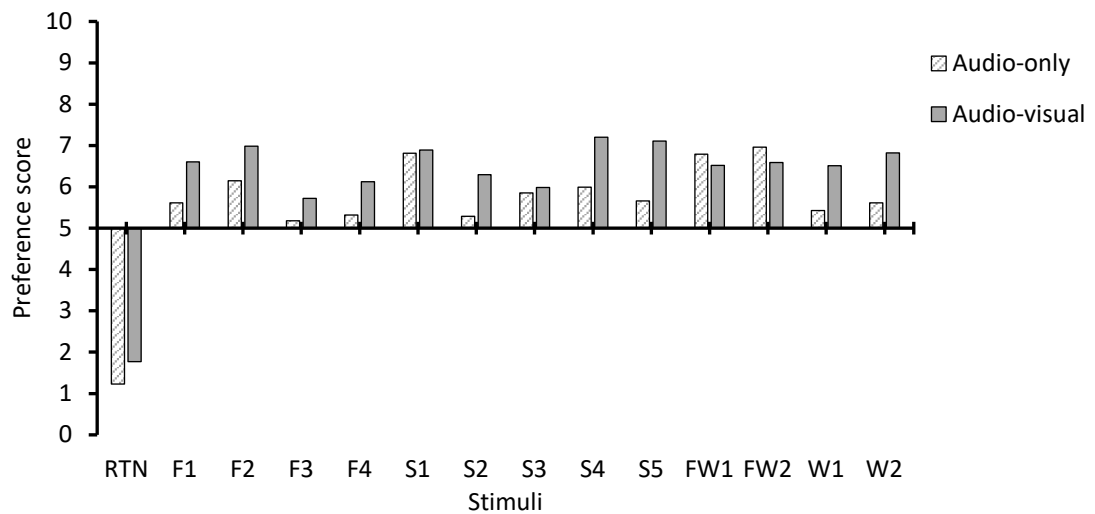


Figure 2.25 Audio-only and audio-visual preference scores of water sounds combined with road traffic noise at an L_{Aeq} of 55 dBA (Jeon *et al.*, 2012). Refer to Figure 2.21 for visual representations of the acronyms (Fig. reproduced).

Presenting visual stimuli with audio materials in Galbrun and Calarco's study (2014) affected the preference ratings of water sounds in comparison to that of the audio-only condition. The differences in preference ratings between the audio-only and audio-visual conditions for ten water sounds are shown in Figure 2.26. Adding the visual stimuli to the water sounds resulted in higher preference scores for four water sounds, namely, the natural stream (ST), the four-step cascade (CA), the saw-tooth edge waterfall (SEW), and the plain edge waterfall (PEW). Whilst, the visual stimuli resulted in lower preference ratings, compared to the audio-only scores, for the rest of the water sounds. However, the changes were significant only for the natural stream (ST) and the narrow jet fountain (NJT). In the case of the natural stream sound, the addition of the visual stimulus significantly increased preference scores compared to the audio-only condition, while for the narrow jet fountain, the opposite was true. Since paired comparisons were used in this study, these results do not necessarily mean that some visual stimuli are detrimental, they merely show that some features benefited more than others from the visual stimuli, confirming results reported by Jeon *et al.* (2012).

The study suggests that adding a stimulus, either visual or auditory, changes the preference and perception of people to water sounds, but the change tends to be statistically insignificant. In addition, the study highlights the interdependence of the audio-only and the audio-visual stimuli, and suggests that equal attention should be paid to the design of both stimuli.

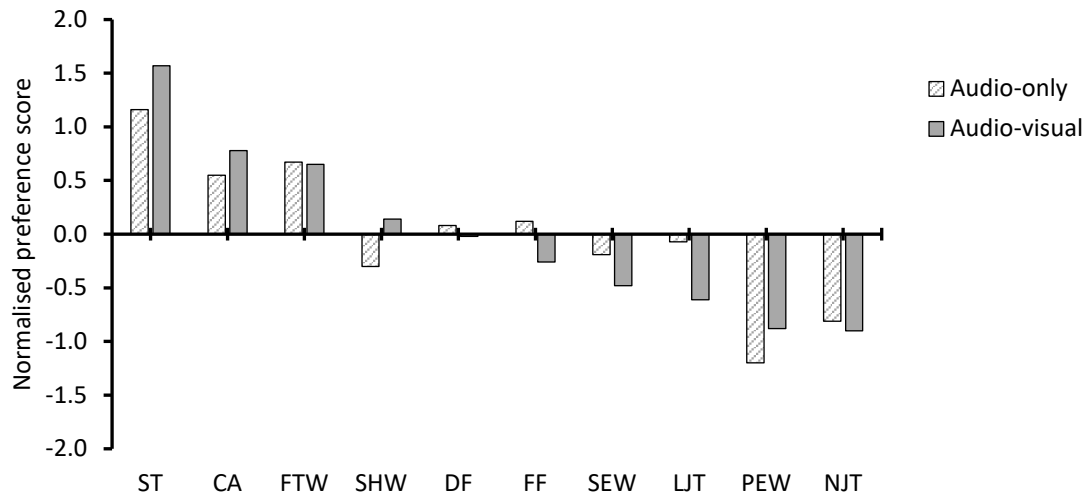


Figure 2.26 Preferred water features from audio-only, and audio-visual tests: normalised preference values as a function of water features (Galbrun and Calarco, 2014) (refer to Table 2.2 for definitions of acronyms) (Fig. reproduced).

In a study on office spaces, people were exposed to four restoration sessions that each lasted seven minutes, after they had performed different cognitive tasks (Jahncke *et al.*, 2011a). The restoration sessions were watching a video clip of a river while listening to its sound, listening to the river sound without watching the video clip, sitting in silence, and listening to office noise.

After the restoration period, participants who had seen the video clip (which included the water sound) rated themselves as being more energetic and less exhausted compared to participants who continued listening to the office noise, or to the river sound alone. Besides, participants who were exposed to the office noise during the restoration session rated themselves as being less motivated when compared to participants who watched the video clip or listened to the river sounds. Higher levels of negative ratings were observed by participants who either listened to the river sound or office noise, in comparison to the other two restoration sessions (river video clip and silence). These results illustrate the potential positive effects of the visual stimulus and support what was claimed by Galbrun and Calarco (2014).

2.13 Single-number quantities

Virjonen *et al.* (2009) examined 16 different open-plan offices in their study, which aimed at developing a single-number quantity that could represent the room acoustic parameters in open-plan offices. The method they proposed relied on measuring the spatial attenuation of the A-weighted sound pressure level of speech ($D_{2,s}$), and the spatial reduction of the STI. The concept of $D_{2,s}$ had previously been adopted in many studies to relate to speech privacy in open-plan offices (Pirn, 1971; Warnock, 1973; West and Parkin, 1975, 1978). The downside of $D_{2,s}$ was that it did not take into consideration the effect of speech masking, therefore, the need for another single-number quantity based on the spatial reduction of the STI was apparent. This new single-number quantity was introduced by Virjonen *et al.* (2009) and was named the *distraction distance*, r_D . The distraction distance is defined as the distance from the speaker at which the STI falls below 0.50. This threshold was based on Hongisto's (2005) model in which an STI value of less than 0.50 was recommended to reduce the detrimental effect of irrelevant speech on cognitive performance. Virjonen *et al.* (2009) also introduced the privacy distance, r_P , which is defined as the distance from the speaker at which the STI falls below 0.20. Within r_P , private and confidential conversations can be carried out. Both r_D and r_P can easily be converted to distraction and privacy areas and if the average density of workers is known, the number of distracted workers can also be estimated. Alongside the $D_{2,s}$ and r_D , Virjonen *et al.* (2009) also measured the sound pressure level of speech at 4 m from the speaker, termed $L_{p,S,4m}$.

The single-number quantities of the 16 offices that were tested by Virjonen *et al.* (2009) were dramatically different. The variations in $D_{2,s}$ values were significantly large, ranging from 4 to 12 dB. The variations in both r_D and $L_{p,S,4m}$ between the office spaces were also large, ranging from 5 to 18 m for r_D and 43 to 54 dB for $L_{p,S,4m}$. Virjonen *et al.* (2009) found that none of the mentioned single-number quantities could alone represent the acoustic parameters of open-plan offices, therefore, they recommended a simultaneous use of $D_{2,s}$, $L_{p,S,4m}$, and r_D to represent the acoustic characteristics of open-plan offices. A similar recommendation was previously made by Virjonen *et al.* (2007) stating that good speech privacy could not be achieved without thinking about all room acoustic components. For instance, speech can travel far from the speaker if the nearby walls and screens reflect the sound, which in turn results in a high value of the $L_{p,S,4m}$, even if the attenuation, $D_{2,s}$, is large. When $D_{2,s}$ is large and $L_{p,S,4m}$ is small (due to highly absorbent ceilings, walls and screens which reduce first-order reflections),

speech may still reach distant listeners if the background noise is low (below 35 dBA i.e., high r_D).

Introducing a masking sound will decrease the STI level, however, it cannot alone guarantee a significant reduction in r_D , if attenuation is small. Besides, r_D cannot be used alone, as small values of r_D can easily be achieved, by using high levels of masking sound and reverberant spaces, which would create a new noise problem. Therefore, attention should be paid to each of the three quantities and the need for using the three quantities all together is evident.

Virjonen *et al.* (2009) classified open-plan offices into 4 categories according to their acoustic properties, as shown in Table 2.3. Class A corresponds to the highest acoustical quality while Class D represents poor acoustics. An open-plan office can have an A in $D_{2,s}$ and $L_{p,S,4m}$, yet a C in r_D .

Table 2.3 Acoustic classifications and target values of open-plan offices (Virjonen *et al.*, 2009).

Class	$D_{2,s}$ (dB)	$L_{p,S,4m}$ (dB)	r_D (m)
A	>11	<48	<5
B	9 to 11	48 to 51	5 to 8
C	7 to 9	51 to 54	8 to 11
D	<7	>54	>11

Virjonen *et al.* (2009) concluded that the reverberation time should not be used as a design criterion in open-plan offices as it represents the temporal attenuation of sound instead of the spatial attenuation. One of the open-plan offices tested by Virjonen *et al.* (2009) had a reverberation time of 0.32 s, yet the value of $D_{2,s}$ was as modest as 6 dB, owing to using low partition screens. Whilst, high values of $D_{2,s}$ were observed even in offices where the reverberation time was higher than 0.5 s.

The single-number quantities proposed by Virjonen *et al.* (2009) were adopted in BS EN ISO 3382-3 (2012) to represent the standardized method according to which room acoustic parameters in open-plan offices should be measured. Detailed information on the measurement procedure of the single-number quantities is given in Chapter 8.

The relationship between the change in the single-number quantities, and the perceived noise disturbance has been examined in a recent study (Haapakangas *et al.*, 2017). The study revealed that changes in the distraction distance largely explain changes in the overall noise perception of work spaces. An increase in the distraction distance was found

to correlate positively with the increase in disturbance by noise, whereas $L_{p,S,4m}$ and $D_{2,S}$ were poor estimators of the perceived noise annoyance. The study showed that an increase of 1 m in the distraction distance can increase the likelihood of recording highly annoyed responses by 9% to 14%. The ratio does not increase with the distance in a linear way. For changes in the distance greater than one metre, the study stated that these ratios can be raised to the power of the distance, to estimate the likelihood of increase of highly annoyed responses. For instance, for an increase of five metres in the distraction distance, the rate of the highly annoyed responses is expected to rise by 54% to 93%².

Research has shown that $D_{2,S}$ and $L_{A,S,4m}$ can be predicted with an accuracy of ± 2 dB and ± 3 dB, respectively. For the prediction of the privacy distance r_D , a variation of ± 4 m should be expected, which is considerable, especially when converted to the distraction area (Keränen and Hongisto, 2013). Therefore, the prediction method was not adopted in the current study.

To conclude, three main strategies should be adopted to improve speech privacy in open-plan offices. These strategies are (Virjonen *et al.*, 2007, 2009):

- increasing absorption near workstations and avoiding reflecting walls and ceilings to reduce first order reflections and eventually decrease $L_{p,S,4m}$;
- using higher and more absorbent screens, which act as sound barriers and increase attenuation, $D_{2,S}$ and;
- introducing a background masking sound, which decreases the signal-to-noise ratio and consequently reduces the STI, which in turn reduces r_D .

The first two points are well understood and related to the room acoustics of the space, while the third one needs more investigation as there are many types of sounds that could be used as speech masking sounds, and very few of them have been investigated.

2.14 Discussion

Bringing water into the built environment would likely add restorative qualities that are normally found in the natural environments, to which water belongs. Water features have successfully been used in parks and plazas to create a pleasant landscape and soundscape. Water sounds are highly appraised by people and preferred over man-made noises such

² Increase in the highly annoyed response = $(1.09)^5$ to $(1.14)^5 = 1.54$ to 1.93, hence 54% to 93% increase in the likelihood of recording highly annoyed responses.

as construction noise and irrelevant speech (Jeon *et al.*, 2010; Jahncke *et al.*, 2016). They have also been proven effective in masking road traffic noise (Watts *et al.*, 2009; Galbrun and Ali, 2013; Galbrun and Calarco, 2014).

Despite the above scientific evidence, water has no strong presence in indoor built environments, possibly due to size restrictions. Carefully designed water features that would take the size limitation and the soundscape perception into account, could overcome issues arising from installing a water feature in indoor spaces, and thus bring the inherent positive qualities of water into these spaces.

The literature identified noise as a major problem in open-plan offices, and irrelevant speech, in particular, has been rated as the most annoying source of noise. Speech masking sounds have been used effectively in past studies to tackle issues arising from the exposure to irrelevant speech (Hongisto, 2005, 2008; Venetjoki *et al.*, 2006; Virjonen *et al.*, 2007; Haapakangas, Haka, *et al.*, 2008; Haapakangas *et al.*, 2011; Liebl *et al.*, 2016). The recommended speech masking sound is generally a pink noise which is filtered to have a spectrum close to that of human speech. Recent studies have shown that natural sounds, especially water generated sounds can also be used as speech masking sounds, yet the results are somehow contradictory. Haapakangas *et al.* (2011) and Keus van de Poll *et al.* (2015) found water sounds to be superior to the filtered pink noise, both in terms of improving cognitive performance and subjective workload. Whilst, Hongisto *et al.* (2017) found filtered pink noise to be preferred over four water generated sounds. The water sounds in these studies were all selected based on their spectral properties, in a way that the water sounds had spectra close to the spectrum of human speech, in line with recommendations given by Veitch *et al.* (2002). However, the latter recommendations were purely based on artificial masking sounds and do not necessarily hold true for natural sounds. Natural sounds have semantic and evocative effects which are not present in artificial sounds. These effects have been proven to greatly affect the perception of natural sounds and their preference levels (Watts *et al.*, 2009; Galbrun and Calarco, 2014). Besides, water sounds that tend to be preferred might not have a spectrum similar to the masked noise (Watts *et al.*, 2009; Galbrun and Ali, 2013). Therefore, identifying the types of water sounds that would be preferred to mask irrelevant speech in open-plan offices is the first step towards providing a more comprehensive answer to the question whether water sounds are appropriate speech maskers. The latter is what this research is aiming at.

Numerous studies have shown that intelligible speech has a detrimental effect on cognitive performance, and masking sounds can reduce this effect. However, the lack of a standardised way to measure performance and due to the fact that different tasks require different cognitive processes, it is difficult to quantify how much improvement in cognitive performance should be expected from installing a masking sound. The model proposed by Hongisto (2005) suggests that masking sounds should aim at reducing the STI of irrelevant speech to below 0.50, to result in a noticeable improvement in cognitive performance. Other studies report even lower values of the STI such as 0.45 (Ebissou *et al.*, 2015) and 0.34 (Jahncke *et al.*, 2013). Hence a masking system that reduces the intelligibility of irrelevant speech from an STI of 0.80 to 0.55, is unlikely to have any beneficial effect on cognitive performance, while the same rate of drop from STI 0.50 to 0.25 would likely result in a significant improvement. Having said that, people often exert more effort to maintain their performance at a constant level, at the cost of increasing the subjective workload (Haapakangas *et al.*, 2011; Jahncke *et al.*, 2011a). In some studies where no significant improvement in the cognitive performance was observed for certain tasks, speech masking sounds had a significant positive effect on the subjective satisfaction of people with the acoustic environment (Haapakangas *et al.*, 2014; Ebissou *et al.*, 2015). Therefore, installing a masking sound in an open-plan office would likely increase either the cognitive performance, the subjective satisfaction or both, although the extent of the increase is not clear.

Due to the vast number of water sounds that could be examined in this research, an extensive literature review was carried out in view of reducing their number through identifying the acoustic and psychoacoustic properties as well as the semantic and evocative effects of water sounds that have been preferred in previous research. This would later help excluding water sounds that are often disliked, which would save time and cost without compromising on the quality of the findings of this research. The literature suggests that water sounds generated from small to medium-sized water features have higher levels of middle-high frequency contents (500 Hz to 16 kHz), while, speech spectrum is characterised by having its peak at middle-low frequencies (125 Hz to 500 Hz). Only waterfalls with a relatively large flow rate are capable of producing high levels of low-frequency sounds that could energetically mask speech. However, waterfall sounds have repeatedly been reported as being disliked by people (Watts *et al.*, 2009; Galbrun and Ali, 2013; Rådsten-Ekman *et al.*, 2013; Galbrun and Calarco, 2014). Furthermore, waterfall sounds have been shown to be difficult to identify (Galbrun and

Calarco, 2014), which would increase the likelihood of them being confused with the noise that they are meant to mask (e.g., irrelevant speech) according to the *target-masker confusion* theory (Durlach *et al.*, 2003; Watson, 2005). Accordingly, attempting to mask speech with waterfall sounds or water sounds which have spectra close to that of human speech might not be an effective approach. On the other hand, sounds made by natural streams and fountains tend to be preferred which would make them potential speech masking sounds (Galbrun and Ali, 2013; Galbrun and Calarco, 2014). Water sounds evocative of man-made sounds tend to be disliked while natural and refreshing water sounds improve people's perception of their acoustic environment (Watts *et al.*, 2009; Galbrun and Calarco, 2014). These descriptive recommendations suggest that waterfall sounds should be avoided while natural stream and fountain sounds should be given further considerations (see Chapter 5).

Apart from the above semantic descriptors, the literature does not provide a clear objective measure by which the preference level of water sounds could be explained. Sharpness and temporal variations (L_{A10} - L_{A90}) have been shown to have good correlations with preference levels (Galbrun and Ali, 2013). Additionally, the sound interference level (SIL) which is the arithmetical mean of the sound pressure level at 500, 1k and 2k Hz, positively correlated with tranquillity ratings in Watts *et al.* (2009). Veitch *et al.* (2002) proposed a new measure, Low-High Frequency A-weighted Level Difference (L_{AeqLow} - $L_{AeqHigh}$), which has been found to correlate well with acoustic satisfaction in open-plan offices (Veitch *et al.*, 2002; Hongisto *et al.*, 2015). The value of L_{AeqLow} - $L_{AeqHigh}$ is obtained by subtracting the A-weighted level of the high-frequency sounds (1 kHz to 8 kHz) from the A-weighted level of the low-frequency sounds (16 Hz to 500 Hz) (Veitch *et al.*, 2002). These measures might provide an objective way by which the preference levels of water sounds could be further explained in the current study.

The visual stimuli of water generated sounds have been shown to play a key role in people's perception of waterscapes. An appropriate visual stimulus accompanying a water sound is likely to result in positive changes in people's perception. Within the context of an open-plan office, this would practically mean installing a water feature in the space instead of playing its sound from a set of speakers. From the soundscape's perspective, installing a water feature is not solely about providing speech privacy; it is also about increasing the aesthetic value of that space. Attractiveness and novelty (interestingness) are considered two crucial aspects of the aesthetic evaluation (Oostendorp and Berlyne, 1978), and carefully designed water features are likely to have

both elements. A water feature could also act as a visual landmark in the space where it is installed. Landmarks often increase the *attractiveness* and *interestingness* of the spaces (Karmanov and Hamel, 2008). Research has shown that water sounds tend to be the first noticeable sounds by people in a space, i.e., create a soundmark (Yang and Kang, 2005). Therefore, installing a water feature in an open-plan office would likely add a landmark as well as a soundmark to the space.

Furthermore, the context in which a sound is played is of a paramount importance. People cannot decide appropriately whether a sound is desirable unless they are aware of the context in which the sound is played (Schafer, 1994). A loud rock music may be perceived as being pleasant during a night party, but it would be extremely annoying in a library. Accordingly, if a water sound were to be played back using a few hidden speakers, an ambiguous context would probably be created in which people might not identify the sound source and thus, become confused. On the ground of the above discussion, the use of a real water feature in this research is justified instead of simply playing its sound from a set of loudspeakers.

2.15 Conclusions

This chapter provided the necessary knowledge that would allow this PhD research to be conducted. Natural environments and the psychological and physiological benefits associated with the exposure to these environments were presented. It was also shown that natural environments have qualities that could be added to built environments through using natural elements such as water.

The chapter highlighted the issues arising from working in open-plan offices. Noise was identified as a major source of annoyance in open-plan offices and irrelevant speech was shown to be the most annoying source of noise. Numerous studies highlighted that irrelevant speech has a detrimental effect on cognitive performance and subjective workload. These detrimental effects can be reduced through using an appropriate masking system, yet the extent of the reduction was difficult to be accurately estimated. The improvement in cognitive performance and subjective workload associated with installing a masking system was shown to be dependent on the cognitive task as well as the masking sounds. Tasks which rely on the short-term memory were found to be more affected by intelligible speech than other tasks, and hence, masking sounds would likely be more beneficial for these tasks. It was also shown that the behaviour of performance

decrement could be predicted from the reduction in the STI associated with installing a masking system.

The masking sounds that had been used in previous studies were covered, which included white noise, pink noise, filtered pink noise, babble, and music as well as natural sounds. Natural sounds, especially water sounds, were shown to be worthy of being used as speech masking sounds. However, due to the lack of recommendations and guidance specific to natural sounds, the selection of water sounds in previous research was based on recommendations which were initially made for artificially generated sounds. This has led to contradicting results with respect to whether water sounds are superior to the recommended artificial masking sounds. Due to this knowledge gap, preference and perception ratings of waterscapes used in urban soundscape studies were discussed in view of providing a better understanding of the types of water sounds that tend to be preferred by people. The literature suggested that stream sounds and fountain sounds tend to be preferred while waterfall sounds tend to be disliked and should be avoided. In addition, evocative and semantic indicators of water sounds were found to be important in determining their preference levels. Water sounds that were perceived as being natural, relaxing, refreshing, and familiar were preferred to water sounds that were perceived as being manmade and rough. In terms of the impact materials, combinations of boulders and water were preferred over water falling over cavities and hard surfaces. Some acoustic and psychoacoustic properties of water sounds were found to be predictable from the knowledge of their design factors. Studies also showed that visual stimulus is a key factor in determining preferences of water sounds. Broadly speaking, the visual stimuli increased the preference of water sounds, yet not all water sounds benefited equally from their visual stimulus.

Lastly, single number qualities were covered that have been proposed to be used as a standardised way in representing room acoustic parameters in open-plan offices. These single-number quantities should be used simultaneously to draw comprehensive conclusions on the room acoustic qualities of open-plan offices. They can also be used in quantifying the benefits associated with making acoustic improvements such as installing a masking system.

The chapter identified key studies and provided necessary information that would allow this research to be carried out. A summary of the main findings presented in this chapter is listed below.

Exposure to natural environment

- Natural environments have the ability to improve well-being and provide restoration from stress or attentional fatigue (Staats *et al.*, 2003).
- Exposure to natural environments is associated with improved negative mood state, improved cognitive functioning and physiological sign of stress reduction (Ulrich, 1983; Ulrich *et al.*, 1991; Kaplan, 1995).
- A well-designed and attractive urban environment could be as stress-reducing and mood-enhancing as an attractive natural environment (Karmanov and Hamel, 2008).

Water features

- Water sounds are believed to be generated as a result of vibrating bubbles and impact sounds caused by water droplets falling on a surface of water (Franz, 1959).
- The frequency of sound generated by water bubbles is correlated with the size of the bubbles (Leighton, 1994).
- Large bubbles break up to form smaller bubbles which act as several individual sound sources (Leighton, 1994).
- Water offers a pleasant contrast to the rigidity and stability of the other architectural elements (Lehrman, 1980).
- Few natural movements are as graceful and attractive as water movements to humans (Burmil *et al.*, 1999).
- People are willing to pay more for houses and hotel rooms overlooking water (Luttik, 2000; Lange and Schaeffer, 2001).
- Urban parks containing water features are perceived as being more restorative than others with no water features (Nordh *et al.*, 2009).
- People prefer spaces that contain *pleasure* and *arousal factors*, and water can provide both factors (Galindo and Corraliza, 2000; Faggi *et al.*, 2013).
- Adding water to the built environment can significantly increase both preference and affective reaction ratings (White *et al.*, 2010).

Noise masking

- Sound masking happens when a tone renders another tone inaudible or less audible (Long 2014).
- Sound masking can happen in two forms, energetic masking, and information masking (Moore, 1995; Durlach *et al.*, 2003).
- The acoustic space is the volume of space in which the sound of a sound source can be heard (Schafer, 1994).
- Acoustic spaces in open-plan offices often exceed their corresponding visual boundaries.

- The congruence and consistency between the acoustic space and its visual boundaries become stronger in the presence of an appropriate masking system.

Human Speech

- Speech patterns can be divided into the audible spectrum and the modulation spectrum (Jacob, 2001).
- Both audible and modulation spectra are characterised by having higher levels of middle-frequency sounds (Steeneken and Houtgast, 1980; Houtgast and Steeneken, 1985).
- Speech intelligibility is affected by the background noise level and the reverberation time (Bradley *et al.*, 1999b; Bistafa and Bradley, 2000).
- The speech transmission index (STI) can be used to objectively measure speech intelligibility (IEC 60268-16, 2011).

Open-plan offices and irrelevant speech

- Open-plan offices increase workers' dissatisfaction and cognitive workload and reduce cognitive performance (De Croon *et al.*, 2005; Hongisto, 2005; Haapakangas *et al.*, 2014; Liebl *et al.*, 2016).
- Noise, especially irrelevant speech, causes the highest level of annoyance in open-plan offices (Hongisto, 2005, 2008; Venetjoki *et al.*, 2006; Virjonen *et al.*, 2007; Haapakangas, Haka, *et al.*, 2008; Haapakangas *et al.*, 2011; Liebl *et al.*, 2016).
- Speech masking is beneficial in reducing the detrimental effect of irrelevant speech in open-plan offices (Veitch *et al.*, 2002; Hongisto, 2005; Haapakangas *et al.*, 2017).
- Room acoustic treatments and speech masking systems should be used simultaneously to achieve optimum speech privacy in open-plan offices (Virjonen *et al.*, 2007).
- The recommended masking spectrum that keeps a balance between speech privacy and comfort level is a pink noise whose sound pressure level reduces 5 dB (Veitch *et al.*, 2002) or 7 dB per octave band (Hongisto *et al.*, 2015).
- The recommended masking sound level in an open-plan office is 45 dBA (Veitch *et al.*, 2002).
- Water generated sounds can outperform some artificial masking sounds, both in terms of increasing subjective satisfaction as well as the cognitive performance in a serial recall task (Haapakangas *et al.*, 2011; Keus van de Poll *et al.*, 2015).
- Using natural sounds as speech masking sounds can restore performance level back to the performance in the absence of speech (quietness) (Jahncke *et al.*, 2016).
- Filtered pink noise resulted in a higher subjective satisfaction than water sound in one study (Hongisto *et al.*, 2017).

Speech intelligibility and cognitive performance

- The error rate in some cognitive tasks can increase by up to 41% because of perfectly intelligible speech (Hongisto, 2005).
- Irrelevant speech can impair short-term memory tasks such as the serial recall task as well as more complex tasks such as proofreading, recalling prior knowledge and logical reasoning. (Landström *et al.*, 2002; Venetjoki *et al.*, 2006; Haka *et al.*, 2009; Smith-Jackson and Klein, 2009; Jahncke *et al.*, 2013, 2016).
- People are believed to exert more effort to maintain their performance at a constant level (Haapakangas *et al.*, 2011; Jahncke *et al.*, 2011a; Ebissou *et al.*, 2015).
- The effect of the sound level of irrelevant speech on cognitive performance is believed to be marginal in comparison to its intelligibility level (Schlittmeier *et al.*, 2008; Jahncke *et al.*, 2011a).
- Performance decrement can be estimated as a function of the STI (Hongisto, 2005).
- Up to an STI of 0.20, cognitive performance remains unaffected by speech intelligibility (Hongisto, 2005; Jahncke *et al.*, 2013).
- At an STI of 0.50, nearly maximum performance impairment has occurred (Hongisto, 2005).
- The *irrelevant speech effect* and the *changing-state hypothesis* are two theories explaining how speech intelligibility interfere with cognitive performance (Salamé and Baddeley, 1989; Jones *et al.*, 1992).
- Tasks which require processing of order information (e.g., serial recall task), tend to be sensitive to unsteady noises such as irrelevant speech (Perham *et al.*, 2009).
- Tasks that require semantic processing (e.g. reading a text, and proofreading), tend to be more impaired by the meaning of irrelevant speech (Marsh *et al.*, 2008; Marsh and Jones, 2010).

Preference of water sounds

- Preferred water sounds do not necessarily have spectra similar to the masked sound (Watts *et al.*, 2009; Galbrun and Ali, 2013).
- Water sounds with higher frequency contents are associated with water falling over hard surfaces (Yang, 2005).
- Water falling onto a cavity produces higher levels of low-frequency sounds (Watts *et al.*, 2009).
- Impact materials play a key role in determining the acoustic and psychoacoustic properties of water sounds (Galbrun and Ali, 2013).
- Water sounds made by water falling over small boulders or a combination of boulders tend to be preferred while water sound produced as a result of water falling over cavities tend to be disliked (Watts *et al.*, 2009; Galbrun and Ali, 2013).

- The sound generated by water features such as streams, fountains, waterfalls and cascades tend to be dominated by middle and high-frequency sounds (Watts *et al.*, 2009; Jeon *et al.*, 2012; Galbrun and Ali, 2013).
- The sound pressure level and loudness of water sounds generated from small to medium-sized water features increase logarithmically with the flow rate (Galbrun and Ali, 2013).
- Stream sounds tend to be preferred over fountain sounds which are in turn preferred to waterfall sounds (Galbrun and Ali, 2013; Rådsten-Ekman *et al.*, 2013; Galbrun and Calarco, 2014).
- The spectral characteristics of water sounds are not the only factor affecting their preference level. Semantic and evocative indicators play a key role in determining the preference levels of water sounds (Watts *et al.*, 2009; Galbrun and Ali, 2013; Galbrun and Calarco, 2014).
- Visual stimuli increase preference scores of water sounds (Watts *et al.*, 2009; Jeon *et al.*, 2012; Galbrun and Calarco, 2014).
- Water sounds perceived as being natural are preferred over water sounds perceived as man-made sounds (Watts *et al.*, 2009; Galbrun and Ali, 2013; Galbrun and Calarco, 2014).
- Among psychoacoustic metrics, sharpness tends to correlate with preference scores of water sounds (Watts *et al.*, 2009; Jeon *et al.*, 2012; Galbrun and Ali, 2013).
- Larger temporal variations are associated with higher preference scores (Watts *et al.*, 2009; Galbrun and Ali, 2013).
- Psychoacoustic metrics such as sharpness, roughness and the temporal variations do not necessarily represent their perceptual descriptors (Galbrun and Calarco, 2014).
- Natural stream sounds tend to be easily identified while fountains and waterfall sounds might be confused with the background noise (Galbrun and Calarco, 2014).
- Congruence between acoustic and visual stimuli is likely to increase preference scores of water sounds (Carles *et al.*, 1992).
- Water sounds benefit differently from their visual stimuli (Jeon *et al.*, 2012; Galbrun and Calarco, 2014).
- Equal attention should be given to audio and visual designs of water features (Galbrun and Calarco, 2014).

Speech privacy metrics

- The rate of spatial decay of the A-weighted sound pressure level of speech per distance doubling ($D_{2,s}$), the sound pressure level of speech at 4 m from the speaker ($L_{p,S,4m}$), and the distraction distance, r_D , are single-number quantities

representing the acoustic parameters in open-plan offices (BS EN ISO 3382-3, 2012).

- The distraction distance, r_D , is the distance from a speaker at which the STI falls below 0.50 (Virjonen *et al.*, 2009).
- Annoyance caused by irrelevant speech at a distance farther than the distraction distance starts to reduce rapidly (Hongisto, 2005; Virjonen *et al.*, 2009).
- The privacy distance, r_P , is the distance from a speaker at which the STI falls below 0.20 (Virjonen *et al.*, 2009).
- Within the privacy distance, private and confidential conversations can be carried out (Virjonen *et al.*, 2009).
- All single-number quantities should be used simultaneously to provide a comprehensive representation of room acoustic parameters in open-plan offices (Virjonen *et al.*, 2009; BS EN ISO 3382-3, 2012).
- The reverberation time should not be used as a design criterion in open-plan offices (Virjonen *et al.*, 2009).

CHAPTER 3: METHODOLOGY AND STATISTICAL ANALYSIS

3.1 Introduction

This chapter presents a general overview of methodologies adopted across the different experiments of the study. Given the varying nature of experiments carried out in this research, this chapter only provides a brief explanation of the methodologies used at various stages of the research. Detailed descriptions of the methodologies are given in their respective chapters. The chapter starts by describing the audio materials (i.e., water sounds and speech recording) used throughout the study, before providing an overview of the methodologies used in five experiments carried out in this research towards meeting its aim and objectives. Statistical models used to analyse the results are then discussed, followed by conclusions.

3.2 Water sounds

Overall, six types of water sounds were used in this study. These included small sized water features that were designed to be installed in outdoor settings as well as indoor environments. The water features were designed and constructed in the laboratory by Galbrun and Ali (2013) and are representative of a wide range of water features that could be used to mask noise. The water features differed in design as well as flow rate and impact materials. The original study examined the effectiveness of these water sounds in masking road traffic noise.

The water features were originally classified into 3 categories, namely, waterfalls, fountains with upward jets, and streams. Galbrun and Ali (2013) suggested that water sounds categorised as waterfalls tend to be disliked by people. Another study (Galbrun and Calarco, 2014) confirmed the same finding. On this ground, it was decided to exclude water sounds categorised as waterfalls in the current study. Furthermore, the natural shallow stream sound, which was highly preferred in previous research (Galbrun and Ali, 2013; Galbrun and Calarco, 2014) was also excluded, owing to the fact that it is not practical to have a stream in an open-plan office. In the original study, water was preferred as an impact material, while hard impact surfaces tended to be disliked by participants.

Thus, one more water feature was excluded as a result of using a hard surface as impact material. Six water features remained in the pool of data after excluding the above water features. These were all considered as small in size and practical to be used in an open-plan office. The water features that were selected were, a cascade with four steps (CA), a dome fountain (DF), a foam fountain (FF), a fountain with 37 upwards jets (FTW), a narrow jet (NJT) and a large jet (LJT). The design properties and acoustic/psychoacoustic characteristics of these water features are given in Table 3.1.

Table 3.1 Acoustic and psychoacoustic parameters of the water sounds normalised at 55 dBA (Galbrun and Ali, 2013).

Sound code	Water feature	Impact material	Fl.	$L_{A\bar{x}-x}$	L_{C-A}	Sh.	Ro.	Pi.
CA	Cascade (4 steps)	Stone (pebbles)	15	-1.30	1.20	2.21	0.10	0.05
DF	Dome fountain	Water	30	0.31	1.60	1.96	0.07	0.14
FF	Foam fountain	Stones <i>and</i> Boulder	30	-0.25	2.30	1.91	0.09	0.05
FTW	Fountain (37 jets)	Water	30	-0.90	1.40	2.21	0.07	0.10
LJT	Large jet	Water	15	4.94	4.90	1.73	0.28	0.08
NJT	Narrow jet	Water	15	-0.96	1.90	2.09	0.19	0.07

Fl. = Flow rate (l/min). $L_{A\bar{x}-x} = L_{A10} - L_{A90}$ (dB). $L_{C-A} = L_{Ceq} - L_{Aeq}$ (dB). Sh. = Sharpness (acum).

Ro. = Roughness (asper). Pi. = Pitch strength

3.2.1 Test structure

The water features described above were generated using a test structure built in the Building Services laboratory of Heriot-Watt University and configured to meet the design properties of each water feature. The test structure is shown in Figure 3.1.

The structure consisted of a sump tank (2.0 m long \times 1.2 m wide \times 1.2 high) encased in the floor, and a tank (1.5 m long \times 0.5 m wide \times 0.5 m high) fixed to a structural frame at a higher level. Two submersible low noise water pumps were placed in the sump tank and used to circulate water to the upper tank or to the fountain extensions. Sound reflections were minimised through using absorption panels and bass traps around the structure. The original study examined a variety of waterfalls and the tank was crucial for configuring the structure to resemble waterfalls. However, since waterfalls were excluded in this study, the only water feature that used the upper tank was the four-step cascade (CA). The remaining water features were mainly fountains and were created using different fountain extensions attached to the water pumps via a pipe. Photographs of the selected water features are shown in Figure 3.2.

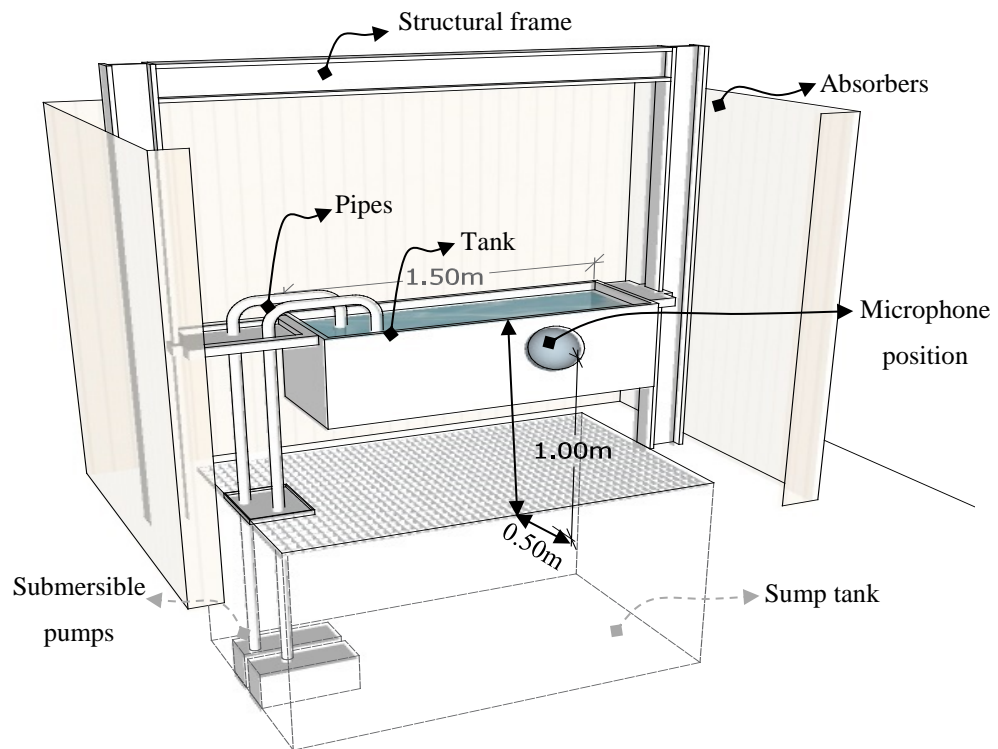


Figure 3.1 Laboratory structure used to test the water generated sounds (Galbrun and Ali, 2013) (Fig. reproduced).

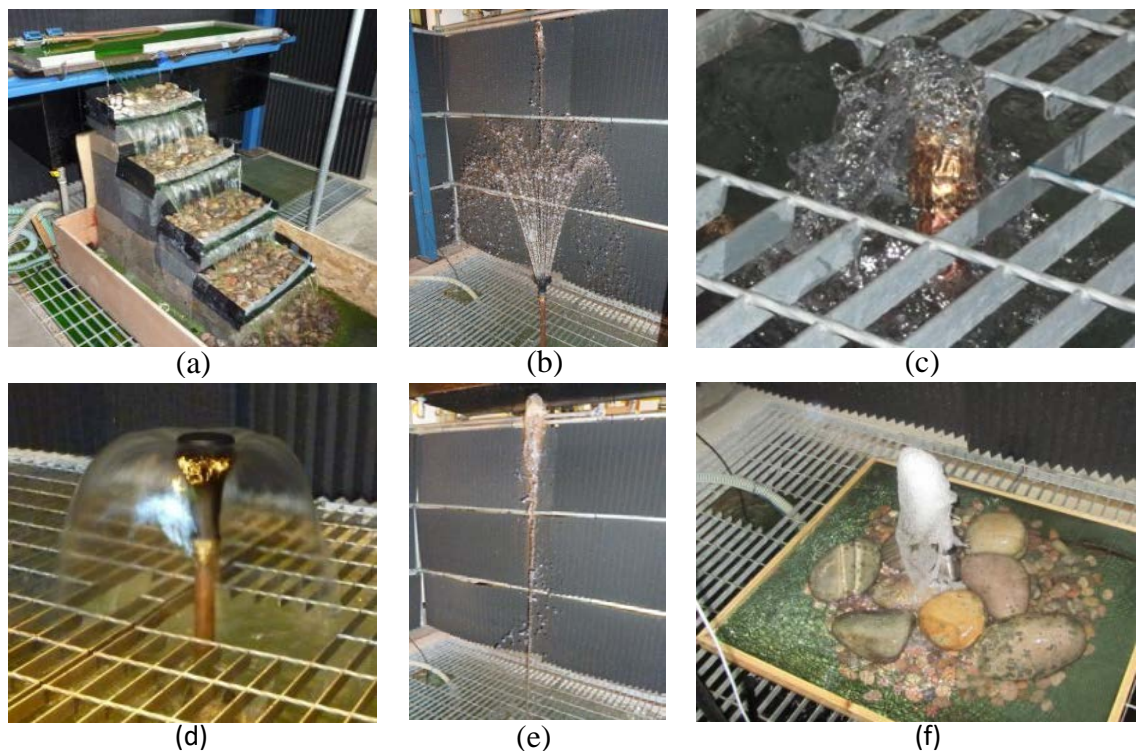


Figure 3.2 Photographs of water features selected in this study, taken at the Building Services laboratory of Heriot-Watt University. (a) Cascade (4 steps). (b) Fountain (37 jets). (c) Large jet. (d) Dome fountain. (e) Narrow jet. (f) Foam fountain. (Galbrun and Ali, 2013).

Audio recording of the water sounds was carried out with a Zoom H4n digital sound recorder connected to Brüel and Kjær Type 4190 ½ microphones, which in turn were attached to a dummy head. The dummy head was placed 0.5 m away from the impact area of the falling water at a height of 1 m above the ground. Given the large size of the laboratory (20 m × 15 m × 7 m), recordings at the dummy head's position was considered as being free from reflected sounds from the boundaries of the space. The result of the audio recordings was a 20-second long binaural recording for each of the water features. A 20-second-long audio recording was considered long enough to cover the operation cycles of the water features (i.e., steady water sounds for all water features tested apart from the large jet, LJT). Seven-second-long audio samples were then extracted from the binaural recordings and were used in the tests run by Galbrun and Ali (2013). Given the steady nature of the sounds, the 7-second period was considered long enough by Galbrun and Ali (2013) to allow for the calculation of the acoustic and psychoacoustic parameters of the water sounds. In the current study, the short audio recordings (i.e., 7-second) were used in the sound level preference test (Chapter 4) and the audio-only and audio-visual preference and perception tests (Chapter 5). The longer recordings (i.e., 20-second) were used in the cognitive performance tests (Chapter 6). More details on the water sounds and their spectral properties are given in their corresponding chapters.

Psychoacoustic parameters such as sharpness, roughness, and pitch strength, were calculated by Ali (2012) using the module *PsySound3* in MATLAB. The following default time steps were used in the calculations: 2 ms for sharpness, 186 ms for roughness, and 10 ms for pitch strength. More specific information on the water features and their sounds can be obtained from Ali (2012).

3.3 Speech recording

The audio recording that has been used as a source of irrelevant speech across different experiments in this study was provided by Dr Jennifer A. Veitch from the National Research Council Canada, Institute for Research in Construction (IRC). The recording had been used in one of the studies carried out by the IRC on the masking of speech in open-plan offices (Veitch *et al.*, 2002).

The original recording consisted of 17 minutes of almost continuous speech of a single female voice speaking at a realistic speech level. The speech comprised one-sided dialogues simulating one side of telephone conversations, represented by the voice of an

actress reading scripts of telephone conversations, in which she called job candidates to arrange for interviews, made internal arrangements for new employees and made personal social calls (Veitch *et al.*, 2002). This study is one of the most comprehensive studies in the area of speech masking in open-plan-office and has been used by numerous studies as a reference in determining the preferred masking sound and sound level.

Seven-second long speech signals were extracted from the speech recording to be used as a source of irrelevant speech in the sound level preference tests (Chapter 4) and audio-only and audio-visual preference and perception test (Chapter 5), to match the length of the water sounds. In doing so, care was taken to extract equally meaningful sentences which contained similar amount of information. The 17-minute long speech recording was used in the task performance tests (Chapter 6) as a source of irrelevant speech. Both the 7-second-long and the 17-minute-long speech recordings were separately calibrated to have appropriate sound levels for the purpose of the tests. More information on the calibration process and the spectral properties of the recordings is provided in their corresponding chapters.

3.4 Methodology

Given the broad area that this study covers, it was important to carry out different experiments with different methodologies, to meet the aims and objectives of the study. Overall, five experiments were carried out, four of which required human participations. A brief overview of the methodologies adopted in each experiment is explained below, and detailed explanations on these methodologies are given in their respective chapters.

3.4.1 Experiment 1: Sound level preference test

The main aim of the experiment was to identify the preferred masking level of the water sounds to be used over irrelevant speech, in open-plan offices. The experiment was also designed to look at the likely effect of the type of water sound, and the intelligibility level of irrelevant speech, on the preferred masking sound level.

Two water sounds, a 4-step cascade (CA), and a 37-jet fountain (FTW) were played at five sound pressure levels, 42, 45, 48, 51, and 54, dBA against irrelevant speech played at a constant sound pressure level of 48 dBA. The speech transmission index (STI) was used as an objective measure of the intelligibility level of irrelevant speech. The STI of the irrelevant speech was altered by adding background noise to it, so that the irrelevant

speech had two STI values, 0.50, and 0.78. This resulted in four test conditions at each masking level:

- FTW played against irrelevant speech with an STI of 0.50;
- FTW played against speech with an STI of 0.78;
- CA played against irrelevant speech with an STI of 0.50; and
- CA played against speech with an STI of 0.78.

Paired comparisons were used. For each condition, five sound pressure levels were tested, which resulted in 10 paired comparisons per condition¹. Hence, the total number of paired comparisons was 40 (10×4). The test was carried out in the anechoic chamber of the acoustic laboratory at Heriot-Watt University, where participants were invited to listen to all pairs of sounds through headphones, and state their preference. Detailed explanations of this experiment are given in Chapter 4.

3.4.2 Experiment 2: Audio-only and audio-visual preference and perception tests

The aim of this experiment was to identify the type of water sound that would be preferred by people to be used as a speech masking sound, and whether people's preferences were different between audio-only and audio-visual conditions. The experiment also examined the change in people's perception of their sound environment, after masking irrelevant speech by a water sound. Six water sounds (Table 3.1) were used and their sound pressure levels were fixed at the level determined by Experiment 1. These were all played against the irrelevant speech which had a constant $L_{Aeq, 7s}$ of 48 dB, and a constant STI of 0.78. The reason for not selecting the STI of 0.50 mentioned in Experiment 1, is given in Chapter 5. The experiment was divided into four main parts:

- Audio-only preference test: Tested the audio-only (no animation) preference of the water sound. Paired comparisons were used which compared all water sounds to each other. This resulted in 15 paired comparisons.
- Audio-visual preference test: Tested the audio-visual (with animation) preference of the water feature. The test was similar to the audio-only preference test, but the audio materials were accompanied by visual materials. The test included 15 paired comparisons.
- Audio-only perception test: Tested the change in people's perception of the sound environment after masking irrelevant speech with each water sound.

¹ Number of pairwise comparisons = $\frac{(N-1) \times N}{2}$, where N is the number of variables (i.e., sound pressure level of masking sound, 5).

- Audio-visual perception test: This test was similar to the audio-only perception test, but the audio materials were accompanied by visual materials.

The tests were carried out in the anechoic chamber of the acoustic laboratory at Heriot-Watt University. Audio materials were played through headphones and the visual materials were presented on a screen. The visual materials consisted of six high quality animations of the water features embedded in an image of a furnished open-plan office. Detailed explanations of this experiment are given in Chapter 5.

3.4.3 Experiment 3: Cognitive performance and subjective satisfaction

The aim of this experiment was to examine the effect that masking irrelevant speech has on people's cognitive performance and subjective satisfaction. In light of the findings from Experiment 2, a water sound (FTW), was selected to be used as a speech masking sound. Participants were invited to take part in the experiment. They performed a set of 4 cognitive tasks under two background noise conditions, a speech-only condition and masked speech condition. The speech only condition was created through playing the 17-minute-long speech recording as background noise. The masked speech condition was achieved by playing the FTW sound over the speech recording. Two speakers were used to play the sounds: one used to play the speech recording and the other used to play the water sound. The four tasks were a serial recall task, a one-back task, an information matching task, and a reading comprehension task. Cognitive workload and subjective difficulty of the tasks were measured using questionnaires distributed in both background noise conditions. The experiment was carried out in Room G.33 of the Edwin Chadwick Building at the Edinburgh Campus of Heriot-Watt University. Detailed explanations of this experiment are given in Chapter 6.

3.4.4 Experiment 4 Longer-term exposure to a water sound

The aim of this experiment was to measure the longer-term impact of a water feature on people's perception of their work environment. In accordance with results of the audio-visual preference test in Experiment 2, a water feature was purchased and modified to match the criteria set out by Experiments 1 and 2. This water feature was placed in a small open-plan office (Room 3.16) in the William Arrol Building at the Edinburgh Campus of Heriot-Watt University. A questionnaire measuring different aspects of the work environment, including the sound environment, was distributed before the water feature was placed in the space. After the water feature had been running in the space for 3 weeks, the same questionnaire (with some added questions specific to the water feature) was redistributed and responses given by participants before and after the installation of the

water feature were compared. Detailed explanations of this experiment are given in Chapter 7.

3.4.5 Experiment 5 Objective measures

This experiment aimed at finding the rate of drop in the STI associated with masking speech in open-plan offices using a water sound. The distraction distance, r_D , and the privacy distance, r_P , are two single number quantities recommended by BS EN ISO 3382-3 (2012) to represent room acoustic parameters in open-plan offices. The r_D is the distance at which STI drops below a value of 0.50, and the r_P is the distance at which the STI drops to a value of 0.20. Both r_D and r_P of two open-plan offices were measured before and after adding a water sound (FTW) to the background noise, and the rate of drop in r_D and r_P was recorded. Measurements in this experiment were carried out in accordance with BS EN ISO 3382-3 (2012). The two open-plan offices were room 2.04 in the Edwin Chadwick Building, and the ground floor of the William Arrol Annexe, both located at the Edinburgh Campus of Heriot-Watt University. Detailed explanations of this experiment are given in Chapter 8.

3.5 Statistical analysis

The data was analysed using IBM SPSS 22 for Windows. A wide range of data was collected in this study. The data differed in their statistical properties, which required using different statistical models. The statistical models used in this study can be classified into two categories, parametric tests, and non-parametric tests. The parametric tests are assumption based and require the data to meet a number of assumptions such as additivity and linearity, normality, homoscedasticity/homogeneity of variance, and independence. If parametric tests were run on data that violates these assumptions, then the statistical results (especially the p -value) would be biased, and therefore, less trustworthy (Field, 2013). There are alternatives for some of the parametric tests, known as non-parametric tests, or “assumption free tests”. These tests make fewer assumptions by ranking the data and using the rankings in the statistical analysis, instead of the actual values of scores within the data. For instance, the smallest score in a data set gets a ranking of 1, the second smallest a ranking of 2 and so on. The ranking process overcomes the shape of distribution of scores.

The preference scores as well as subjective satisfaction scores measured using questionnaires were all considered to have violated the assumptions of the parametric tests and thus, non-parametric tests were used in their statistical analyses. Data collected

as part of the task performance tests were considered to have met, or partially met the assumptions, and therefore, parametric tests were used for their statistical analyses.

3.5.1 *The mean*

Throughout this study, the mean, M , has been used as a measure of the central tendency of the data (i.e., where the centre of a frequency distribution lies). The mean is simply the average of all scores within a data set.

3.5.2 *Confidence intervals*

When the sample distribution is normal, confidence intervals are a good way to estimate how representative a sample is of the population from which the sample is drawn. They result in boundaries within which the true mean of the population is believed to fall. Large confidence intervals indicate that the sample is a poor representative of the population, while small intervals are indicative of a good representation of the population by the sample. Throughout this study, the 95% confidence intervals (95% CI) have been adopted and reported alongside the mean values in square brackets []. A 95% confidence interval is an interval constructed such that in 95% of samples, the true value of the population mean will fall within its limits (Field, 2013). The upper and lower boundary of the 95% confidence interval can be calculated using Equation (3.1):

$$\begin{aligned}\text{lower boundary} &= \bar{X} - (1.96 \times SE) \\ \text{upper boundary} &= \bar{X} + (1.96 \times SE)\end{aligned}\tag{3.1}$$

where

\bar{X} is the mean of the sample;
 SE is the standard error of the sample.

The 95% CIs are also useful in graphs, as these allow to visually estimate if the means of two samples are significantly different. If the confidence intervals overlap substantially, then it is very unlikely to find any statistically significant differences. Whilst if the confidence intervals of the two samples do not overlap, or moderately overlap, then the chances are high to find statistically significant differences. As a result, the 95% CIs are provided in most graphs used in this study to enhance the interpretation of the results, visually.

3.5.3 *Bootstrap Confidence intervals*

The confidence intervals assume a normal sampling distribution, and their values can significantly be compromised when the assumption of normality is violated. Bootstrap

(Efron and Tibshirani, 1993) is a technique used to estimate confidence intervals that are robust to the violation of the assumption of normality. SPSS can calculate 95% bootstrap confidence intervals using two methods, a percentile confidence interval, and a bias corrected and accelerated confidence interval, 95% BCa CI. The latter method is more accurate (Efron and Tibshirani, 1993) and has therefore been adopted in this study to calculate robust CIs for non-normally distributed samples, across the various experiments.

3.5.4 Outliers

Outliers are scores which are significantly different from the rest of the data. They can bias the value of the mean and the confidence intervals which in turn results in erroneous statistical results. Outliers are identified mathematically in this study as values which are 2.5 times greater/smaller than the mean values. Boxplots have also been used to visually spot potential outliers. Outliers are excluded from the statistical analyses in this study.

3.5.5 Null hypothesis significance testing and the *p*-value

This research is based on a number of hypotheses (or research questions). These hypotheses mainly assume that an effect of an experiment is present, and are known scientifically as *experimental hypothesis*. There is another type of hypothesis which is known as *null hypothesis*, and is the opposite of the experimental hypothesis. The experimental hypothesis cannot be proved using statistics, but the null hypothesis can be rejected to provide more support to the experimental hypothesis. The procedure of rejecting/accepting the null hypothesis starts by assuming that the null hypothesis is true, then fitting a statistical model to the data that represents the experimental hypothesis. To determine how well the model fits the data, the probability, denoted as *p*-value, of obtaining the model if the null hypothesis was true, is calculated. Among the scholars, a *p*-value = .05 is accepted as a cut-off point, and therefore, any *p*-value smaller than .05 results in rejecting the null hypothesis and increasing the confidence that the experimental hypothesis is true (Field, 2013). Across this study the *p*-value < .05 has been adopted as a criterion to reject the null hypothesis.

3.5.6 Comparing two means

The parametric test of comparing two means and testing whether they are significantly different is known as a *t*-test. In all cases where such comparisons were needed in this study, the assumptions of parametric tests could not be assumed, and therefore, the non-parametric versions of the *t*-test have been used.

When different participants were used, known as *independent measure design*, such as comparing scores between male and female participants, the Mann-Whitney test was used to carry out the statistical analysis. The test statistic for the Mann-Whitney test is denoted by U .

When the same participants were used, known as *repeated measure design*, such as comparing satisfaction scores in a questionnaire before and after an intervention (e.g., Chapter 7), the Wilcoxon signed-rank test has been adopted to carry out the statistical analysis. The test statistic for Wilcoxon signed-rank test is z .

3.5.7 Comparing several means

The parametric test for comparing several means is known as the *analysis of variance* (ANOVA) which is a generic term used to cover a variety of statistical tests. When different participants are used, the test is known as an independent ANOVA, and when the same participants are used the test is called a repeated ANOVA. The test statistic for ANOVA is F (referred to as F -ratio), which is the ratio of systematic variance to unsystematic variance in an experimental study (Field, 2013).

ANOVA can be extended to include one or more continuous variables that can predict the dependent (outcome) variable, but are not part of the experimental manipulation. For instance, the noise sensitivity score (NoiseQ) of participants in Chapter 6 has been used as a covariate to help improve the ANOVA analysis. When such covariates are included, the test is called the analysis of covariance, ANCOVA.

When more than one independent variable is included in the design of the ANOVA (or ANCOVA), it is called factorial ANOVA (or ANCOVA) where there are more than one independent variable predicting the output or dependent variable. For example, in Chapter 6, participants' response accuracy (RA) is a dependent variable, and the gender, nationality and the background noise conditions are 3 independent variables all predicting the response accuracy.

The concept of ANCOVA can be further extended by including two or more dependent variables simultaneously, in the analysis. In such cases, the analysis is known as the Multi Variate Analysis of Covariance (MANCOVA). Most of statistical analyses in Chapter 6 are based on MANCOVA, as there are two dependent variables, namely, response accuracy (RA) and reaction time (RT). Running separate ANCOVA on each RA and RT would inflate the Type I error (incorrect rejection of a true null hypothesis), and hence, MANCOVA was used. In addition, MANCOVA allows for examining the relationship

between the dependent variables, something which ANCOVA is not capable of. The importance of using MANCOVA over ANCOVA becomes apparent in Chapter 6.

The non-parametric versions of ANOVA are known as Kruskal-Wallis test for independent design (test statistic is denoted by H), and Friedman's ANOVA for repeated measure design (test statistic is denoted by χ^2). These tests are similar in principle to ANOVAs but they use a ranking system instead of using the actual data. Statistical analyses in Chapter 4 and 5 were made using these tests.

3.5.8 Factorial designs

The statistical tests described above can have three different designs:

- Independent design: this type of test is suitable for experiments where different groups of participants are used for different test conditions. For instance, comparing responses from different age groups.
- Repeated measure (related) design: this type of test is suitable for experiments where the same participants are used in all test conditions. Most of the statistical analyses in this study have been carried out using this type of design, as the same participants took part in all test conditions.
- Mixed design: this design is suitable when some of the dependent variables have been measured using the same participants, whilst other dependent variables have been measured using different groups of participants. An example would be the statistical analysis used in Chapter 6 to compare people's cognitive performance in two background noise conditions, while simultaneously, comparing cognitive performance between males and females. This scenario requires a mixed analysis as comparing cognitive performance under the two background noise conditions is a repeated measure design, while comparing females' and males' performances is an independent design.

3.5.9 Post hoc procedure

The statistical tests discussed above indicate whether experimental groups are significantly different (i.e., groups come from different populations). They do not identify which group(s) is/are different from the other(s). Paired comparisons can be used to compare all the possible pairs; however, this procedure inflates the rate of the Type I error (α). Post hoc tests are designed to control this inflation, by adjusting the p -value. There are several versions of Post hoc tests, and the one that has been used throughout this study

is the Benjamini-Hochberg procedure which controls the expected proportion of falsely rejected hypothesis i.e., the false discovery rate (Benjamini and Hochberg, 1995). This method of adjusting α is statistically more powerful than the Bonferroni correction which calls for controlling the familywise error rate (Benjamini and Hochberg, 1995).

3.5.10 Correlation analysis

Correlations are used to examine the relationship between two variables. Correlation analyses in this study have been carried out using Pearson's correlation coefficient, r , when the assumption of normality could be assumed, and the Spearman's correlation coefficient, r_s , when normal distribution could not be assumed. Both r and r_s result in a value between -1 (a perfect negative correlation) and 1 (a perfect positive correlation). A value of 0 means that there is no correlation between the two variables.

3.5.11 Effect size

The effect size², r , is an objective and standardised measure of the magnitude of the observed effect. The value of r varies between 0 (no effect) and 1 (perfect effect). Unlike the p -value, the effect size has no cut-off point (e.g., $p < .05$), and its value is not affected by the sample size. Therefore, r values are comparable across different experiments and studies. Cohen (1988) suggests the following classification for the size of r :

- $r = .10$ (small effect): the effect explains 1% of the total variance
- $r = .30$ (medium effect): the effect explains 9% of the total variance
- $r = .50$ (large effect): the effect explains 25% of the total variance.

The American Psychological Association (APA) recommends reporting effect sizes. Throughout this thesis, wherever appropriate, effect sizes have been reported to show the magnitude of the effect of the experimental manipulations, and to help making informed decisions about the results. Attempts have been made to further quantify the classifications (i.e., small, medium and large) within the context of each experiment carried out as part of this research (e.g., what would a medium effect size mean in terms of improving peoples' performance?). The calculation process of the effect size differs according to the type of statistical test used. Details about the calculations are given in their respective chapters.

² There are other types of effect size, but only r , which essentially is the Pearson's correlation coefficient, has been used in this study, because it is easier to interpret and can be classified into small, medium and large.

3.5.12 Reliability analysis

The reliability analysis is a way to gauge how consistent a measure (e.g., a questionnaire) is, in reflecting the construct that it is measuring (Field, 2013). Cronbach's alpha, α , is a common reliability measure, which allows to examine whether individual items (or sets of items) produce results which are consistent with the overall questionnaire (i.e., whether individual items in a questionnaire measure the same construct that the questionnaire is designed for). Cronbach's alpha can have a value between 0 to 1. A value of .7 to .8 is commonly accepted as representative of a reliable scale. However, a value below .7 does not necessarily indicate an unreliable scale (Field, 2013). Hence, Cronbach's alpha has been used in the current research to measure the reliability of scales adopted in questionnaires in some experiments (Chapter 8).

3.5.13 Categorical analysis (chi-square test)

The chi-square test has been used to analyse categorical data (perception tests in Chapter 5). The chi-square test compares the observed frequency of selecting a variable, to its expected frequency. Within the context of this research, the test allows for examining whether an option has been significantly more chosen by participants, in comparison to another option. The test statistic for the chi-square test is denoted by χ^2 .

3.6 Conclusions

In this chapter, the methodologies adopted to carry out five experiments were briefly presented. The water sounds used throughout these experiments and the structure used to generate the sounds, were illustrated alongside the acoustic and psychoacoustic characteristics of the water sounds. The speech recording used as a source of irrelevant speech in the experiments was described. The statistical models used to carry out the statistical analyses were also explained. In-depth explanations of the methodologies of each experiment are given in their respective chapters.

CHAPTER 4: SOUND LEVEL PREFERENCE TEST

4.1 Introduction

Identifying the preferred sound pressure level (SPL) of the water sounds was considered to be a good starting point towards meeting the objectives of this research. There are a few guidelines available in the literature regarding the appropriate sound level of masking sounds, but they are restricted to artificial types of masking sound such as pink noise and white noise. The preferred SPL of a masking sound might also be affected by the level of speech intelligibility as well as the type of the masking sound. These two factors seem to have been overlooked in the currently available literature, by recommending a constant SPL of masking sound, regardless of the intelligibility level of the background noise and the type of masking sound. In this chapter, a range of SPLs of two water sounds is tested under two speech intelligibility conditions; high intelligibility level represented by an STI of 0.78 and a lower intelligibility level represented by an STI of 0.50. The methodology adopted to run the experiment is explained with descriptions of the procedures and participants who took part in the experiment. Results are presented and analysed followed by a critical discussion of the results achieved. The chapter comes to an end by providing conclusions of the main finding of the experiment.

4.2 Water sounds

Two water sounds were selected as speech masking sounds. The water sounds were taken from a previous study carried out at Heriot-Watt University by Galbrun and Ali (2013) on the acoustical and perceptual assessment of water sounds and their use over road traffic noise. In the study, numerous small to medium-sized water features were examined in controlled conditions using different design factors such as the flow rate, height of falling water, waterfall's edge design and impact materials. The study was discussed in detail in Chapter 2, but due to its relevance to the current study, a brief summary of its findings is provided here. Overall 12 different water features were tested. The water features were categorised under three groups; Waterfall, Fountain, and Stream. Water sounds with the highest preference scores were a natural stream (ST), a 37-jet fountain (FTW), a large jet fountain (LJT), and a 4-step cascade (CA). The water sounds were further analysed in a similar study by Galbrun and Calarco (2014) and the highly rated water sounds were found to be ST, FTW, and CA, while LJT was poorly rated. On the above grounds, CA

and FTW were chosen to be considered in this stage of the study. ST was excluded despite being very favourably rated, as it would be impractical to install a stream-like water feature in an open-plan office. Also, LJT which was the third highly preferred water sound in one study was excluded due to its low score in the later study. Since this experiment is dedicated to identifying the preferred SPL of the water sounds, it was decided to only include two types of water features which were highly preferred in previous research. Including more than two types of water sounds would have resulted in considerably longer tests which could have negatively affected the accuracy of the responses. The octave-band spectra of the two water sounds are shown in Figure 4.1.

4.3 Speech recording

The high quality 17-min-long speech recording described in Section 3.3 of Chapter 3 was used as a source of irrelevant speech. Similar to previous studies (Galbrun and Ali, 2013; Galbrun and Calarco, 2014), the water sounds were only 7 seconds long; therefore, the speech recording had to be cut into 7-second-long recordings too. Forty 1-sentence speech recordings were extracted from the original speech recording. The arithmetical average spectrum of all 40 1-sentence speech signals is shown in Figure 4.1. Each of the 1-sentence speech recordings was 7 seconds long, to match the length of the water sounds, and were individually calibrated to have an equivalent sound pressure level, $L_{Aeq,7s}$, of 48 dB. According to laboratory experiments conducted by Virjonen *et al.* (2007) the sound pressure level of normal effort speech at a neighbouring workstation in an open-plan office varies between an L_{Aeq} of 39 dB and 55 dB. Similar measurements were made by Hongisto *et al.* (2004) and Virjonen *et al.* (2009). However, the level of speech never fell below 45 dBA in the nearest workstation. In addition, a study has shown that when the background speech level exceeds 48 dBA, it becomes too loud for a masking system to work effectively (Jahncke *et al.*, 2013). The study from which the speech recording is provided (Veitch *et al.*, 2002), set the speech level at 54.5 dBA at 1 m away from the speaker. This setting resulted in a speech level ranging from 41.2 to 44.4 dBA across workstations. The reference level of 54.5 dBA is around 3 dB lower than the speech level of 57.4 dBA recommended by BS EN ISO 3382-3 (2012) to be used as a reference speech level in open-plan offices. Using the recommended reference speech level would have increased the speech level range across workstations in the study of Veitch *et al.* (2002) to be between 44.2 and 47.4 dBA. The speech level of 48 dBA used in the current study is therefore justifiable on the grounds of the above field evidence. This level represents

the highest level of speech at which a masking system could be beneficial. An improvement in speech privacy at this speech level would ensure a greater benefit at lower speech levels.

4.4 Potential masking levels

The identification of the preferred masking level required playing the water sounds at various sound pressure levels. Veitch *et al.* (2002) recommended the level of masking sound not to exceed 48 dBA. Similar recommendations were given by Haapakangas *et al.* (2011) who suggested that the A-weighted sound pressure level of a masking sound should not be lower than 40 dBA, and should not exceed 45 dBA. These empirical findings imply that the sound pressure level of a masking sound (i.e., water sounds) would be somewhere between 40 and 48 dBA. Based on these findings, five different sound pressure levels, namely, 42, 45, 48, 51, 54 dBA were considered to be appropriate to be examined. The higher sound pressure levels (i.e., 51 and 54 dBA) exceeded the highest recommended masking level of 48 dBA, yet they were considered in the current study as previous recommendations were based on artificial masking sounds (e.g., pink noise), whilst the current study uses natural sounds and people might have a higher tolerance towards hearing natural sounds at a higher level. The highest masking level was set at 54 dBA as any masking level exceeding this level would probably be considered too loud, and thus, might cause annoyance itself. The 3-dB difference between any two consecutive levels was seen appropriate to be used as differences smaller than 3 dB would have made it increasingly difficult for participants to distinguish between the levels, and would have increased the risk of guess work. Hence, the sound pressure levels of the water sounds that were examined in this study were 42, 45, 48, 51, and 54 dBA. This range covers the same SPL of speech (i.e., 48 dBA) as well as two levels below and two levels above it.

4.5 Intelligibility of background speech

Background noise in open-plan offices often includes irrelevant speech as well as other background noises such as noise coming from photocopying machines, paperwork and typing, to name a few. These noise sources affect the intelligibility of irrelevant speech by reducing its signal-to-noise ratio. This experiment therefore aimed at identifying the preferred sound pressure level of the water sounds as a function of the speech intelligibility of irrelevant speech. The speech transmission index, STI, was adopted as an objective measure to represent the intelligibility level of irrelevant speech. The STI of the speech recording was altered to have two levels, 0.78 and 0.50. The higher STI (i.e.,

0.78) represents an “excellent” speech intelligibility level. It was achieved by using the dry speech recording plus some reverberant field. Digital audio processing was used to add an average reverberation time of 0.5 seconds. The 0.5 seconds reverberation time was assumed by taking the averages of reverberation times in 15 open-plan offices reported in a previous study (Virjonen *et al.*, 2009)¹. The lower STI (i.e., 0.50) represents a “fair” speech intelligibility level and was achieved by using the dry speech recording, 0.5 seconds reverberation time, and some background noise. The background noise had an $L_{Aeq,7s}$ of 43 dB (i.e., SNR +5) and was added to the speech signal to reduce its intelligibility level. A previously recorded high-quality background noise was selected from the catalogue of “audiosparx.com” after being subjectively reviewed in terms of audio quality, sample length and speech content. The background noise of a busy open-plan office was selected, which had a steady sound level and did not contain any intelligible parts in order not to interfere with the irrelevant speech used in the test. The background noise included footsteps noise, typing and paperwork noises, as well as distant unintelligible speech. The octave-band spectrum of the background noise is shown in Figure 4.1. No lower STI values were included in the study because it would have made the need of using a masking system questionable, as the speech intelligibility would already be low.

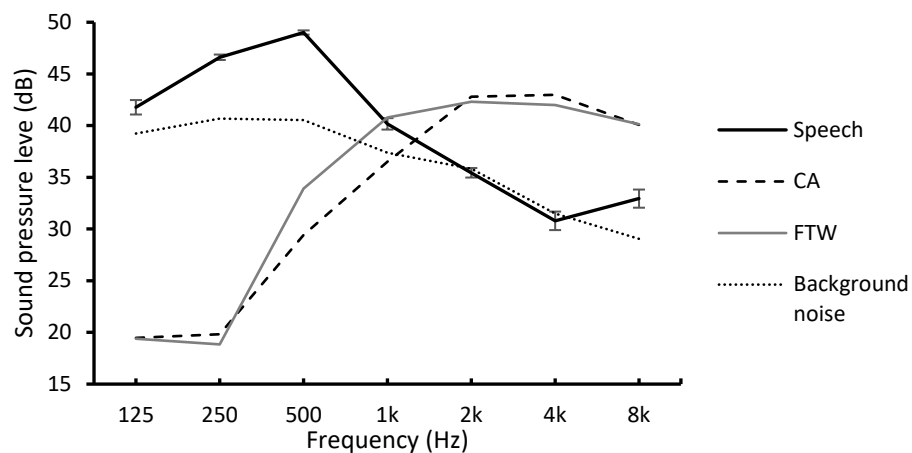


Figure 4.1 Octave-band spectra of irrelevant speech (48 dBA), CA (48 dBA), FTW (48 dBA), and background noise (43 dBA). Error bars on the speech spectrum represent 95% confidence intervals.

Since the speech recording consisted of 40 7-second-long speech signals, the spectrum of each signal varied slightly (shown by the 95% confidence intervals in Figure 4.1). Hence,

¹ Virjonen *et al.* (2009) measured the reverberation time in 16 open-plan offices, but one office had a reverberation time which was considerably longer than others (1.37 s). This office was excluded in calculating the average reverberation time used in the current study.

the STI also varied across the 7-second-long speech recordings. However, the variation was small and did not exceed the mean value ± 0.03 , which is below the just noticeable difference in STI (Bradley et al., 1999a). The calculation process of the STI is explained in Section 2.8 of Chapter 2.

4.6 Participants

Thirty-nine participants (17 males, and 22 females), aged between 23 yr and 48 yr ($M = 30.90$ yr, $SD = 5.84$ yr) took part in the sound level preference experiment. They were staff members and postgraduate students of Heriot-Watt University. Two participants reported having tinnitus, and 9 more participants did not perform well in the consistency test (consistent judgments within a 95% confidence interval; see Section 4.9). Therefore, 28 participants (15 males, and 13 females) were retained for further analysis. Out of the retained sample, 9 participants were native English speakers, and 19 were non-native English speakers who spoke English fluently. Due to the unbalanced number in native and non-native groups, no statistical analysis has been run comparing responses given by these two groups. Participants were given a £5-Amazon voucher as a token of gratitude for taking part in the test.

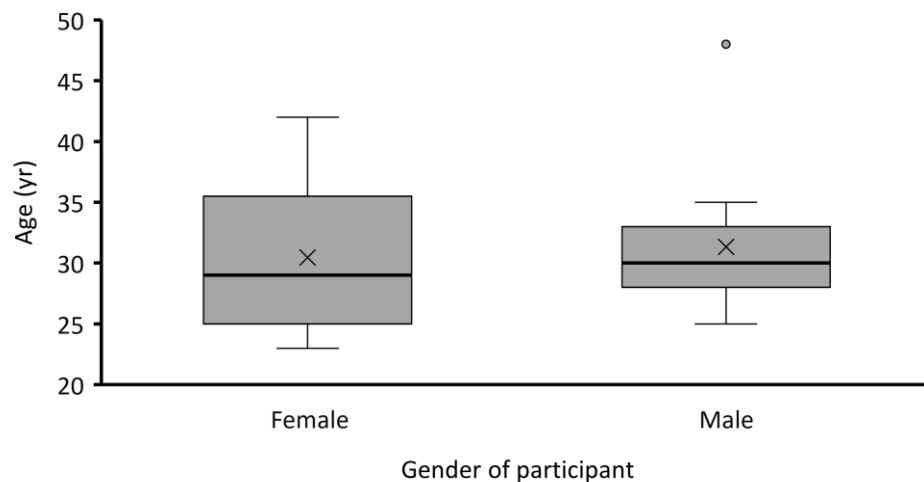


Figure 4.2 Age distribution of participants in the sound level preference test. The tinted boxes represent the interquartile range (i.e., 50% of scores) with the top and bottom of the tinted box representing the upper and lower quartiles, respectively. The bold horizontal lines inside the boxes are medians, and the cross marks (\times) are the means. The small circle represents an unusual case and the whiskers show the top and lowest 25% of scores.

The difference in age distribution of participants between males and females is shown in the boxplots in Figure 4.2. The boxplots show that the average age of males and females

were comparable, but the distribution was more widely dispersed for females. There was one male participant whose age was unusually higher, which might have slightly biased the mean age of male participants.

4.7 Methodology

The two water sounds, (CA and FTW) were played at five sound pressure levels (42, 45, 48, 51, and 54 dBA), against irrelevant speech which had a constant sound pressure level of 48 dBA. The STI of irrelevant speech was also altered to have two levels, 0.78, and 0.50. This combination resulted in four test conditions at each masking level:

- FTW played against irrelevant speech with STI of 0.78 (FTW-STI.78);
- FTW played against irrelevant speech with STI of 0.50 (FTW-STI.50);
- CA played against irrelevant speech with STI of 0.78 (CA-STI.78);
- CA played against irrelevant speech with STI of 0.50 (CA-STI.50).

Paired comparisons were used to compare all masking levels within each test condition. Five sound pressure levels were tested, which resulted in 10 paired comparisons per condition². Hence, the total number of paired comparisons was 40 (10×4). Each paired comparison consisted of a 7-second-long water sound played at a certain SPL over one of the 7-second-long speech signals, followed by the same water sound at a different SPL played over the same speech signal. There was a gap of 1 second between the two audio signals. Five extra paired comparisons were prepared to be used in the practice session.

It is worth mentioning that the combined audio recordings contained a binaural signal (i.e., water sounds), a stereo signal (i.e., the background noise) and a mono signal (i.e., irrelevant speech). These were all combined to make one scene which was reproduced over headphones. Spatial errors are possible when this relatively simple headphone reproduction is used to reproduce such a complex audio signal. In order to minimise unexpected spatial defects, the spatial aspects of the audio reproduction were subjectively tested for spatial errors such as head localisation, front-back reversals, and non-externalised sound fields, and none of these were noticed in the reproduced audio. Furthermore, no participant reported experiencing unexpected spatial effects. Therefore, the audio recordings were assumed to be free from spatial errors.

² Number of pairwise comparisons = $\frac{(N-1) \times N}{2}$, where N is the number of variables (i.e., sound pressure level of masking sound, 5).

4.8 Procedure

The test was carried out in the highly insulated anechoic chamber of the acoustic laboratory at Heriot-Watt University, where participants were invited to listen to all pairs of sounds through headphones, and state their preference.

Participants were seated on an upholstery office chair in the middle of the anechoic chamber in front of a standard office desk. On the desk, there was an evaluation form and a laptop used to play the audio files and present instructions. On the evaluation form, participants were asked to provide basic background information, such as age, gender, and nationality. They were also asked to confirm that they did not have any known hearing difficulties. Detailed instructions were provided on the screen of the laptop, and verbal clarifications were provided when needed. Participants were asked to imagine that they were working in an open-plan office, where they could hear a colleague speaking over the phone from a nearby workstation, as well as a water sound. For each pair of sounds, participants were instructed to tick the sound that they preferred to work in over a long period of time. Given the similarities and the subtle difference between the sound levels in some of the paired comparisons, participants were given a third option which was “no preference”, although they were discouraged from choosing it. Therefore, for each paired comparison, participant would either choose “sound 1”, “sound 2” or “no preference”. In each paired comparison, the sound level that was preferred would get a score of 1 and the other sound level would get a score of zero. If the “no preference” option was chosen, both sounds in that pair would get a score of 0.5. A copy of the evaluation form is provided in Appendix A and the instructions given to participants are provided in Appendix B.

A practice session was provided for the participants to make themselves familiar with the test. This session included five paired comparisons, and participants could repeat the practice session as many times as they pleased. During the practice session, the researcher remained in the room and observed the participants to make sure they understood how the test was performed. Once the participants were confident to carry out the test on their own, the researcher would leave the room and the test would start.

Each participant listened to all 40 paired comparisons in a randomised order, and hence a PowerPoint presentation was prepared for each participant. Each paired comparison was followed by a 3-second silence, during which the participant had to state their preference, before another pair was played. The sounds were set on auto play and after playing 10

pairs of sound, the PowerPoint presentation would stop to allow for the participant to take a short break. The participant then needed to press Enter on the keyboard to play another 10 pairs of sound. The first 10 paired comparisons were repeated at the end of the test to be used as a measure of consistency of responses given by the participants. Participants whose scores were below the lower bound of the confidence interval of the consistency test (see Section 4.9) were removed from the analysis. The test lasted between 40 and 45 minutes, including the short breaks.

Paired comparisons were used in this stage as they have been proven successful in previous soundscape research (Jeon *et al.*, 2010; You *et al.*, 2010; De Coensel *et al.*, 2011; Galbrun and Ali, 2013; Galbrun and Calarco, 2014). Paired comparisons inherently result in ordinal data that makes preference ranking easy. They are preferred to verbal or numerical rating scales because of their simplicity and greater accuracy (Mantiuk *et al.*, 1981). They allow for subjects to choose between only two options which will ensure that some sounds are preferred over others, unlike Likert scales that would allow subjects to give identical scores to different scenarios.

One disadvantage of using paired comparisons is the distribution of data from the statistical point of view, which tends to deviate from a normal distribution. This can make the statistical analysis of the data limited to using non-parametric tests which are arguably less powerful than their parametric counterparts. However, the advantages of obtaining ordinal data from paired comparisons outweigh their disadvantages in the current study.

4.9 Consistency test

To rectify for guess work, a consistency test was embedded in the experiment. The first ten paired comparisons for each participant were repeated at the end of the test, in order to test how consistent their responses were. Ideally, a participant's preference scores in these paired comparisons would be identical to those in the first ten paired comparisons. However, given the subtle differences between the sound levels, some discrepancies would be anticipated. The participant's response in a paired comparison in the consistency test would then be compared to the response of the reference paired comparisons and if the preference was different, then the consistency score would be reduced by 10%. For example, a participant who managed to score seven identical responses out of a total of ten would be given 70% in the consistency test.

After the consistency score of each participant had been measured, the average consistency score and the 95% confidence intervals were calculated. For this test, the average consistency was $M = 75.41$ [70.03, 80.27]. The lower limit of the confidence interval, i.e., 70.03 was then taken as a cut-off point to exclude participants from the analysis whose score was below 70%. Nine participants scored less than this cut-off point and therefore their responses were excluded from the statistical analysis. The following statistical analysis is based on a sample size of $N = 28$.

4.10 Equipment and digital audio processing

The audio files were all played through a pair of Beyerdynamic DT 150 closed headphones which were connected to a laptop (Acer aspire S3 series, model No.: MS2346) through an M-Audio external USB sound card. Microsoft PowerPoint 2013 was used to create the presentations and play the audio files. The water sounds and speech signals were calibrated using a Brüel and Kjær handheld sound analyser, Type 2250. Digital audio processing was carried out using Studio One 3 audio production software (PreSonus Audio Electronics) installed on a Dell XPS L502X personal computer.

4.11 Statistical analysis

The ordinal nature of the data makes non-parametric statistical models more appropriate for the statistical analysis of the data in this chapter. Ordinal data are non-continuous as the variables within the data can take only integral values, which violates the assumption of continuity of variables within data on which most parametric models are founded (Field, 2013).

The data were analysed using IBM SPSS 22 for Windows. Since all participants participated in all test conditions, Friedman's Analysis of Variance (ANOVA) was used to test the likely effect of the speech intelligibility level and the type of masking sounds on the level preferences. The test was also used to test whether the preference scores given to the masking sound levels were different, i.e., whether the preference scores came from the same population or represented different populations. Pairwise follow-up analysis was carried out when Friedman's ANOVA showed a significant difference among the groups or conditions. The p -values were adjusted using the Benjamini-Hochberg procedure which controls the expected proportion of falsely rejected hypothesis i.e., the false discovery rate (Benjamini and Hochberg, 1995). In addition, the effect sizes of the pairwise comparisons were calculated. It is worth mentioning that both the Benjamini-

Hochberg procedure and Effect size calculations are not readily given by SPSS and therefore, Microsoft Excel was used to calculate them. SPSS provides z-scores which can be converted to an effect size, r , using the following equation (Field, 2013):

$$r = \frac{z}{\sqrt{N}} \quad (4.1)$$

where N is the number of observations, which is twice the number of participants (28×2) in this case, since the same participants took part in all test conditions.

Differences in rating between males and females, as well as between different age groups, were tested using the Mann-Whitney test. The Bias-corrected and accelerated bootstrap method, BCa, was used to derive robust 95% confidence intervals, which are reported in square brackets throughout this chapter.

4.12 Results

Twenty-eight participants passed the consistency test and were retained for the analysis of results. Preference scores for each condition, i.e., FTW-STI.50, FTW-STI.78, CA-STI50, and CA-STI78, plus the overall preference scores were normalised to have values between -2 (never preferred) to +2 (always preferred), for easy comparisons. A summary of the normalised preference scores is provided in Figure 4.3.

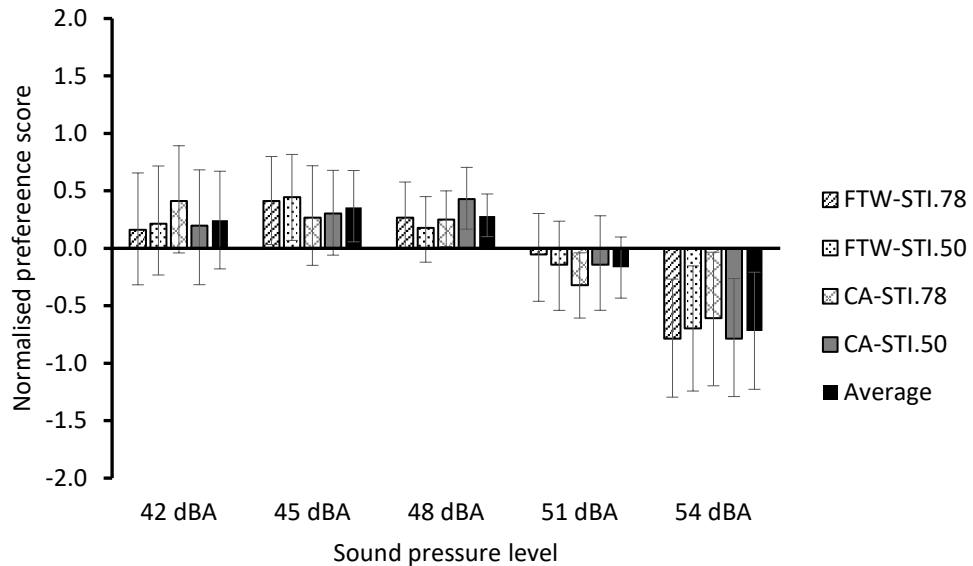


Figure 4.3 Normalised preference scores for the four test conditions alongside the averaged preference scores. Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals.

The “no preference” option was chosen 9% of the time (106 times). Initially, it was decided to give a score of zero for both sound levels included in a pair when the no preference option was selected. Later, it was discovered that the no preference option was not randomly distributed over the pairs of sounds. Generally, participants struggled to state their preference when the difference in the SPL between the two sound levels was 3 dB, and this became more apparent at the lower levels. For example, the pair that compared sound level 42 dBA to 45 dBA, received 28 “no preference” options, while the pair that compared 51 dBA to 54, received only 17 “no preference” options. The no preference trend across all ten paired sounds is shown in Figure 4.4. To avoid bias in results and to overcome this issue, it was decided to give a value of 0.5 to both sounds (i.e., SPL) included in a pair when that pair received a no preference.

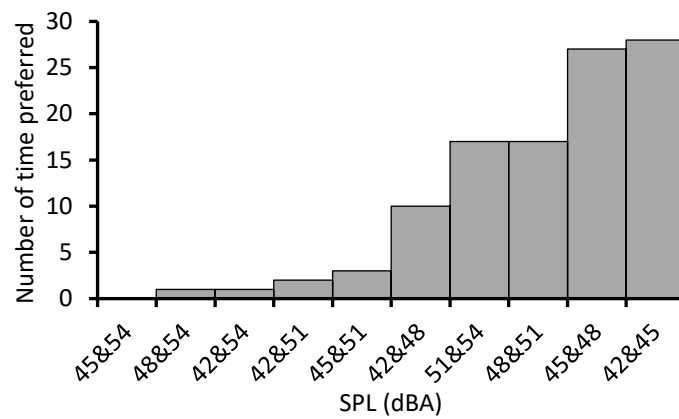


Figure 4.4 Number of times the "no preference" option was chosen for each pair of sounds.

4.12.1 Effect of speech intelligibility and type of water sound on preference scores

For each sound pressure level, 2 properties were altered, the STI of the speech signal (.50 and .078) and the type of water sound (CA and FTW), resulting in four test conditions. The first step in the statistical analysis was to examine if participants' preferences were different in these four test conditions. The statistical analysis revealed that the alteration in the STI and the type of the water sound did not have a significant impact on preference scores at any of the masking levels, 42 dBA ($\chi^2(3) = 2.683$, $p = .443$), 45 dBA ($\chi^2(3) = 0.451$, $p = .929$), 48 dBA ($\chi^2(3) = 1.967$, $p = .579$), 51 dBA ($\chi^2(3) = 0.904$, $p = .824$), and 54 dBA ($\chi^2(3) = 1.739$, $p = .628$). The analysis suggests that at each level, people perceived the four conditions similarly. Therefore, an average score from all four conditions was calculated (solid black bars in Figure 4.3) and retained for further analysis.

4.12.2 Preference of sound pressure level of the water sounds

Bias-corrected and accelerated bootstrap 95% CIs are reported in square brackets. The average normalised preference score at each sound level was calculated and is shown in Figure 4.3. The result revealed that the most preferred masking sound level, irrespective of the type of water sound or the STI of irrelevant speech, was 45 dBA, $M = 0.36$ [0.06, 0.68], followed by 48 dBA, $M = 0.28$ [0.10, 0.47], and 42 dBA, $M = 0.25$ [-0.18, 0.66], respectively. The least preferred masking sound level was 54 dBA, $M = -0.72$ [-1.23, -0.21], followed by 51 dBA, $M = -0.17$ [-0.43, 0.10]. The confidence intervals provide more information regarding the preference scores. For 45 and 48 dBA, the 95% CI remained positive, i.e., did not pass through the zero line, which gives more confidence to the positive preference scores given to 45 and 48 dBA. Similarly, the 95% CI for the level 54 dBA remained negative without passing through zero giving more confidence to the negative preference score. However, for levels 42 and 51 dBA, their 95% CI passed through zero, making the positive score of 42 dBA and the negative score of 51 dBA less reliable. Therefore, it can be concluded that a sound pressure level between 45 and 48 dBA is preferred for a water sound to be used as a speech masker in an open-plan office.

Statistically, preference scores were significantly affected by the sound level of the water sound, $\chi^2(4) = 14.268, p = .007$. Pairwise comparisons with Benjamini-Hochberg adjusted p -values were used to follow up this finding. It appeared that the 54 dBA level was significantly less preferred than 45 dBA ($z = -3.254, p = .010, r = -.435$), 48 dBA ($z = -2.916, p = .020, r = -.390$), and 42 dBA ($z = -2.747, p = .020, r = -.367$). No further statistically significant differences were detected between preference scores of the other sound levels, all $p > .05$.

The gender of participants did not have a significant impact on preference scores at any of the masking levels; 42 dBA ($U = 68.00, z = -1.366, p = .178, r = -.258$), 45 dBA ($U = 67.50, z = -1.386, p = .172, r = -.262$), 48 dBA ($U = 96.50, z = -0.047, p = .972, r = -.008$), 51 dBA ($U = 62.50, z = -1.616, p = .109, r = -.305$), and 54 dBA ($U = 74.50, z = -1.067, p = .296, r = -.202$).

Participants were divided into two age groups, *below 30 yr* ($N = 14$), and *30 yr and above* ($N = 14$). No statistically significant differences in preference were detected between these two groups at any sound level; 42 dBA ($U = 40.00, z = -1.351, p = .186, r = -.255$), 45 dBA ($U = 55.50, z = -0.330, p = .759, r = -.062$), 48 dBA ($U = 40.00, z = -1.362, p =$

.183, $r = -.257$), 51 dBA ($U = 38.00$, $z = -1.482$, $p = .147$, $r = -.280$), 54 dBA ($U = 58.50$, $z = -0.132$, $p = .910$, $r = -.025$).

4.13 Discussion

The experiment resulted in the identification of the preferred sound pressure level of water sounds when used to mask irrelevant speech. The preferred sound level was found to be 45 dBA, which was 3 dB lower than the speech level of 48 dBA used in this study. This confirms the previously recommended range of level of masking sounds (Veitch *et al.*, 2002). It should also be noted that preference scores given to 42 and 48 dBA were not significantly lower than that of 45 dBA, suggesting that these levels can also be advantageous. This range of preferred levels (42 dBA – 48 dBA) allows for some flexibility in designing a masking system, by having higher than ideal levels close to a noise masking source i.e., a water feature, and lower than ideal levels farther from the source. Furthermore, this range seems to be independent of the type of water sound and the intelligibility level of the background speech. The lower STI level of 0.50 may still be considered high enough, which might justify the similarity in preference scores between STI 0.50 and 0.78. Preference scores given by male and female participants were similar. In addition, no significant difference was detected between preference scores given by people from two age groups, below 30 yr and 30 yr and above. More research is needed to attain more conclusive findings regarding whether the preferred level of water sound is affected by the intelligibility level of the background speech.

4.14 Conclusions

In this chapter, the preferred sound pressure level of two water sounds was determined, when used for masking irrelevant speech in open-plan offices. The results showed that people preferred the water sounds to have a sound pressure level of 45 dBA. This corresponded to a masking level 3 dB lower than the sound pressure level of irrelevant speech used in this study (i.e., 48 dBA). In addition, preference scores given to 42 dBA and 48 dBA were not significantly lower than the most preferred level, 45 dBA. This range of preferred levels did not seem to be dependent on the type of water sound, the speech intelligibility level of irrelevant speech, nor the gender of participants. The next chapter builds upon the findings achieved here, and extends the study to examine the audio-only and audio-visual preferences for six water features used to mask irrelevant speech in open-plan offices.

CHAPTER 5: AUDIO-ONLY AND AUDIO-VISUAL PREFERENCES AND PERCEPTION

5.1 Introduction

After the preferred sound pressure level of the water sounds had been identified in Chapter 4, audio-only and audio-visual preference and perception tests were carried out, in view of identifying the type of waterscape that would be preferred as a speech masking sound, while positively affecting people's perception. This chapter presents the steps taken towards carrying out this experiment. Audio-only and audio-visual preferences of six waterscapes are examined, within the context of speech masking in open-plan-offices (preference test). In addition, the likely changes in people's perception, associated with masking irrelevant speech with each of the water sounds are investigated (perception test). The chapter starts by providing a brief description of the water sounds and the speech recording used, followed by descriptions of participants who took part in the experiment. The methodologies adopted in both preference and perception tests are explained in detail, alongside the statistical models used to analyse the data collected. Results are then analysed and the main findings are presented and critically discussed. The chapter comes to an end by providing conclusions of the main findings and a brief overview of the upcoming chapter.

5.2 Water sounds

The water sounds included in this study were taken from a previous work (Galbrun and Ali, 2013), in which a range of water features was examined under controlled conditions. Detailed descriptions of the structure used to produce the water sounds were provided in Chapter 3. Six water sounds were selected to be included in this experiment. The water sounds were a 4-step cascade (CA), a 37-jet fountain (FTW), a dome fountain (DF), a foam fountain (FF), a large jet (LJT), and a narrow jet (NJT).

The spectral properties of the water sounds are shown in Figure 5.1 and Table 5.1. All water sounds were calibrated to an $L_{Aeq,7s}$ of 45 dB in line with findings of the sound level preference test (Chapter 4). Some acoustic and psychoacoustic characteristics of the water sounds are shown in Table 5.2, which includes the Sound Interference Level (SIL), and

the Low-High Frequency A-weighted Level Difference ($L_{AeqLow}-L_{AeqHigh}$). The SIL is the arithmetical mean of sound pressure levels at 500, 1k and 2k Hz. The $L_{AeqLow}-L_{AeqHigh}$ is the difference between the A-weighted sound pressure levels of low (16-500 Hz) and high (1-16 kHz) frequency sounds. These two measures have been reported to correlate well with preference levels (Veitch *et al.*, 2002; Hongisto *et al.*, 2015) and tranquillity levels (Watts *et al.*, 2009) of water sounds.

5.3 Speech recording

Similar to the sound level preference tests (Chapter 4), 42 7-second-long speech signals were extracted from the original 17-minute-long speech recording described in Chapter 3. Eight extra speech samples were prepared to be used in the practice tests, resulting in 50 7-second-long speech signals, in total. Out of the 50 speech signals, 40 samples were identical to those used in the sound level preference test. Each speech signal was calibrated to have an $L_{Aeq,7s}$ of 48 dB. Justifications for selecting this level were given in Chapter 4. The octave-band spectrum of the speech signals is shown in Figure 5.1, and their acoustic properties are shown in Table 5.1. These values were calculated separately for each 7-second-long speech signal and averaged out arithmetically to represent the speech signal. The 95% confidence intervals are provided alongside the averaged values as a measure of variability of the values. The STI of the speech signals was fixed at 0.78 achieved through adding an average 0.5 seconds reverberation time to the dry speech signal, via digital audio processing.

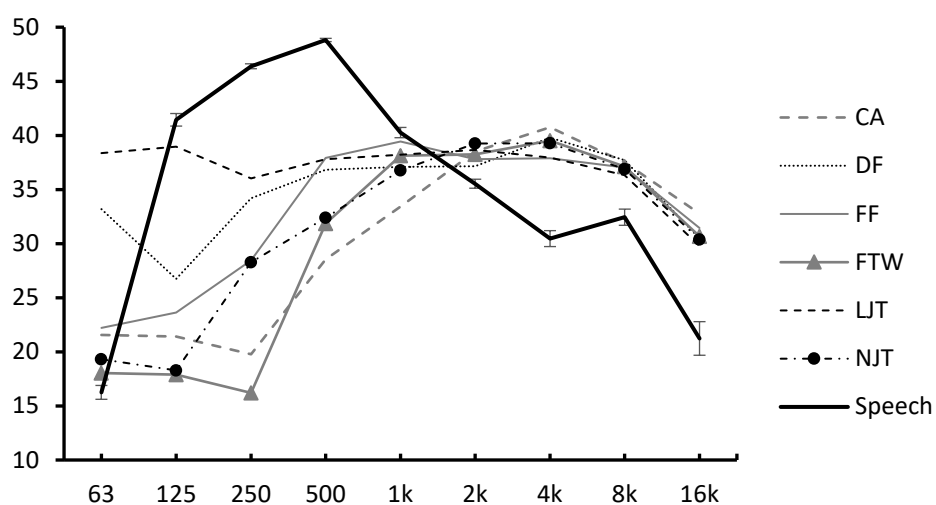


Figure 5.1 Octave-band spectra of six water sounds (45 dBA) and the spectrum of irrelevant speech (48 dBA) used in the audio-only and audio-visual tests. Error bars represent the 95% confidence intervals.

Table 5.1 Octave-band sound pressure levels (dB) of six water sounds as well as the speech signal used in the audio-only and audio-visual preference and perception tests.

Freq. (Hz)	CA	DF	FF	FTW	LJT	NJT	Speech*	95% CIs	
63	21.6	33.2	22.2	18.0	38.4	19.3	16.3	[15.6,	16.9]
125	21.4	26.7	23.6	17.9	39.0	18.3	41.4	[40.9,	42.0]
250	19.8	34.2	28.5	16.2	36.0	28.3	46.4	[46.2,	46.6]
500	28.6	36.8	37.9	31.9	37.8	32.4	48.8	[48.7,	49.0]
1k	33.4	37.1	39.4	38.1	38.2	36.8	40.3	[39.8,	40.7]
2k	38.5	37.1	37.8	38.2	38.7	39.3	35.5	[35.1,	35.9]
4k	40.8	39.8	37.9	39.5	38.0	39.3	30.5	[29.7,	31.2]
8k	37.5	37.7	37.0	37.1	36.3	36.9	32.4	[31.7,	33.2]
16k	32.8	30.5	31.4	30.7	29.8	30.4	21.2	[19.7,	22.8]
$L_{Aeq, 7s}$	45.0	45.0	45.0	45.0	45.0	45.0	48.0	N/A	N/A

*Values for the speech signal are based on the arithmetical mean of corresponding values of 42 7-second-long speech signals. The 95% confidence intervals of the average values are provided in square brackets [lower bound, upper bound].

Table 5.2 Acoustic characteristics of six water sounds as well as the speech signal used in the audio-only and audio-visual preference and perception tests. Units are in dB, unless stated otherwise.

Measures	CA	DF	FF	FTW	LJT	NJT	Speech*	95% CIs	
L_{Aeq}	45.0	45.0	45.0	45.0	45.0	45.0	48.0	N/A	N/A
L_{Ceq}	43.7	45.3	44.8	44.1	49.9	44.0	52.1	[52.0,	52.2]
$L_{Ceq} - L_{Aeq}$	-1.3	0.3	-0.3	-0.9	4.9	-1.0	4.1	[4.0,	4.2]
L_{AFmax}	46.5	47.0	48.7	47.5	53.1	47.3	54.9	[54.5,	55.4]
L_{AFmin}	43.5	43.1	42.2	43.4	39.5	42.9	23.8	[2.2,	25.4]
$L_{AFmax} - L_{AFmin}$	3.0	3.9	6.5	4.1	13.6	4.4	31.1	[29.3,	32.9]
L_{F10}	45.6	45.8	46.1	45.7	47.1	46.0	51.7	[51.4,	51.9]
L_{F90}	44.4	44.2	43.8	44.3	42.2	44.1	33.5	[31.3,	35.6]
$L_{F10} - L_{F90}$	1.2	1.6	2.3	1.4	4.9	1.9	18.2	[15.9,	20.5]
SIL	33.5	37.0	38.4	36.1	38.2	36.1	41.5	[41.3,	41.8]
L_{AeqLow} (16-500 Hz)	25.6	34.3	34.9	28.7	35.6	29.7	46.3	[46.2,	46.5]
$L_{AeqHigh}$ (1k-8k Hz)	45.0	44.6	44.6	44.9	44.4	44.8	42.8	[42.5,	43.1]
$L_{AeqLow} - L_{AeqHigh}$	-19.4	-10.3	-9.7	-16.2	-8.8	-15.2	3.5	[3.0,	3.9]
Sharpness, <i>acum</i>	2.21	1.96	1.91	2.21	1.73	2.09	-	-	-
Roughness, <i>asper</i>	0.10	0.07	0.09	0.07	0.28	0.19	-	-	-
Pitch strength	0.05	0.14	0.05	0.10	0.08	0.07	-	-	-

*Values for the speech signal are based on the arithmetical mean of corresponding values of 42 7-second-long speech signals. The 95% confidence intervals of the average values are provided in square brackets [lower bound, upper bound].

5.4 Participants

Thirty-three participants who reported a normal hearing ability took part in the audio-only and audio-visual tests. Participants were postgraduate students and staff members of Heriot-Watt University. Out of the tested sample, 31 participants (16 males, 15 females) passed a consistency test (consistent judgements within a 95% confidence interval) and were retained for further analysis. The age distribution of retained participants ranged between 24 yr and

60 yr ($M = 36.35$ yr, $SD = 9.32$ yr). Eighteen participants were native English speakers, and the remaining 13 were non-native speakers who spoke English fluently.

The difference in age distribution of participants between males and females is shown in the boxplots in Figure 5.2. The boxplots show that on average, male participants were younger than female participants, and the age distribution was smaller for male participants. There is one male participant whose age was unusually higher, but the mean (the \times mark in Figure 5.2) and the medians (the horizontal line in Figure 5.2) are very close, which suggests that the mean values were not biased by this extreme case (i.e., the small circle in Figure 5.2). Therefore, no participants were excluded based on their age. At the end of the experiment, participants received a 5-pound Amazon voucher, as a token of gratitude for taking part in the experiment.

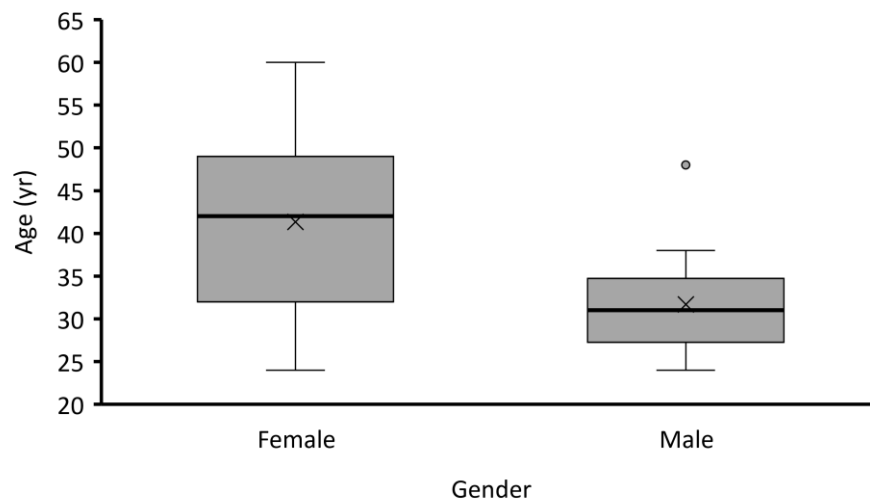


Figure 5.2 Age distribution of participants in audio-only and audio-visual preference and perception tests. The tinted boxes represent the interquartile range (i.e., 50% of scores) with the top and bottom of the tinted box representing the upper and lower quartiles, respectively. The bold horizontal lines inside the boxes are medians, and the cross marks (\times) are the means. The small circle represents an unusual case and the whiskers show the top and lowest 25% of scores.

5.5 Methodology

The experiment was divided into two main sections, a preference test and a perception test. In the preference test, participants stated their preference for the waterscapes, while in the perception test, they rated how their perception of the sound environment changed after irrelevant speech was masked by each of the water sounds. Both tests included audio-only and audio-visual conditions. The preference test identified the water feature(s) that tended to be preferred but no conclusion could be drawn on whether the water sounds

positively contributed to the sound environment (due to the use of paired comparisons, which would become apparent later). The perception test was included to cope with this shortcoming in the preference test.

5.6 Procedure

The tests were carried out in the highly insulated anechoic chamber of the acoustic laboratory at Heriot-Watt University, where participants were invited to listen to all pairs of sounds through headphones and see animations of the water features via a monitor screen. In all tests, participants had to imagine that they were working in an open-plan office where they could hear a water sound and a colleague speaking over the phone at a nearby workstation.

It is worth mentioning that the anechoic chamber did not resemble an open-plan office. However, research has shown that imagination can change people's perception of reality. Berger and Ehrsson (2013) suggest that neuronal signals generated by imagined stimuli can integrate with signals generated by real stimuli of a different sensory modality to create robust multisensory percepts. The study shows that imagining an audio stimulus or a visual stimulus can change people's perception in a similar way to what the real stimulus would do. Hence, in the current study, asking participants to imagine that they were working in an open-plan office might have overcome the fact that the test was carried out in an anechoic chamber instead of a real open-plan office.

Participants were seated on an upholstery office chair in the middle of the anechoic chamber in front of a standard office desk. Underneath the desk, there was a laptop used to play the sounds and the animations. On the desk, there was a monitor screen on which instructions and animations of the water features were presented. An evaluation form was prepared to allow participants to mark their preferences and rate how their perception changed. The evaluation form also required participants to provide basic background information, such as age, gender, and nationality. Participants were also asked to confirm that they did not have any known hearing difficulties. A copy of the evaluation form is provided in Appendix C and the instructions given to participants are provided in Appendix D.

5.6.1 Preference test

The preference test was similar to the level preference test (Chapter 4), but, instead of changing the level of the water sounds, their types were changed. The sound pressure level of the water sounds was fixed at an $L_{Aeq,7s}$ of 45 dB, in line with findings of the sound level preference test. Paired comparisons were adopted to carry out the preference tests. The test was carried out under two conditions, an audio-only condition and an audio-visual condition. In both conditions, six water sounds/water features were compared which resulted in 15 paired comparisons¹ per condition (i.e., 30 paired comparisons in total).

In the audio-only condition, each pair of sounds consisted of a 7-second-long water sound played over a 7-second-long speech signal, followed by another 7-second-long water sound played over the same speech signal. There was a 1-second-long gap between the two sounds. The sounds were played through headphones, and no visual materials were included.

The audio-visual preference test was similar to the audio-only preference test, but visual animations of the water sounds were added to the test. Each pair of sounds consisted of a 7-second-long water sound played over a 7-second-long speech signal, followed by another 7-second-long water sound played over the same speech signal. There was a 1-second-long gap between the two sounds. Each sound was accompanied by its visual animation, presented on the monitor screen in front of the participant.

High-quality realistic animations of the water sounds were prepared and presented to participants, in the audio-visual condition. A photograph of a furnished open-plan office was used as a background image, and the animations of the water features were embedded in the background image. The background image did not include any human figures, to allow for participants to concentrate on the water features and avoid any visual distractions. Still images of the animations are presented in Figure 5.3.

For each paired comparison, participants had to select the water sound/water feature which they preferred working in over a long period of time, and helped them to concentrate. Participants stated their preference by ticking either “Option 1” or “Option 2” on the evaluation form for each paired comparison. Any option that was preferred in a

¹ Number of pairwise comparisons = $\frac{(N-1) \times N}{2}$, where N is the number of variables (i.e., 6 water sounds).

pair of sounds would get a score of 1, and the other option a score of zero. Given the less subtle differences between the water sounds (in comparison to the sound level preference test) the “no preference” option was removed from the evaluation form.



Figure 5.3 Still images from animations of the water features used in the audio-visual condition of the preference and perception tests. (a) Four-step Cascade. (b) Dome fountain. (c) Foam fountain. (d) 37-jet fountain. (e) Large jet. (f) Narrow jet. Background image (Flickr: Space n.d.).

A practice session was provided for the participants to make themselves familiar with the test. This session included 4 paired comparisons (2 audio-only and 2 audio-visual), and participants could repeat the practice session as many times as they pleased. During the practice session, the researcher remained in the room and observed the participants to

make sure they understood how the test was performed. Scores given in the practice session were not included in the analysis of results. Once a participant was confident to carry out the test, the researcher would leave the room and the test would start.

The sequence of paired comparisons was randomised for each participant, but the audio-only condition was always carried out before the audio-visual condition. After each paired comparison, participants had 3 seconds to mark their preference on the evaluation form, before another pair of sounds started to play. Each participant listened/saw all 30 paired comparisons (plus 10 repeated pairs used as a consistency measure; see Section 5.6.3). The sounds and animations were set on auto play. After playing 13 audio-only paired comparisons, the presentation would stop to allow for the participant to take a short break. Then the participant would need to press Enter on the keyboard to play the remaining 12 audio-only paired comparisons (10 of which were repeated pairs for the consistency test), before the presentation stopped again. Next, the participant would need to press Enter one more time, to play the following 15 audio-visual paired comparisons. The test lasted between 30 and 35 minutes, including the short break.

5.6.2 Perception test

After participants completed the preference test, they would call the researcher to come into the test room to provide instructions on the perception test. Similar to the preference test, both audio-only and audio-visual scenarios were examined. In the audio-only condition, participants listened to 7 seconds of background speech, 1 second of silence and 7 seconds of background speech masked with one of the water sounds. Then they were given 5 seconds to rate their perception of the sounds, i.e., “how your perception changed after introducing the water feature”. The scale used by participants to make their evaluation was a 5-point Likert scale and the labels were “much worse, slightly worse, no change, slightly better and much better”. For each water sound, the participant needed to tick one of the labels on the evaluation form until all six water sounds were evaluated.

The process for the audio-visual condition was similar to that described above. The difference was for the first 7 seconds, participants would listen to the background speech while they would see an image of an open-plan office (i.e., without a water feature), followed by 1 second of silence/blank screen, and then the animation of a water feature would appear in the office with its corresponding water sound, before they made their evaluation. This process was repeated for all six water sounds.

A practice session was provided for the participants at the beginning of the perception test to make themselves familiar with the test. This session included 4 waterscape evaluations (2 audio-only and 2 audio-visual), and participants could repeat the practice session as many times as they pleased. During the practice session, the researcher remained in the room and observed the participants to make sure they understood how the test was performed. Scores given in the practice session were not included in the analysis of results. Once the participant was confident to carry out the test, the researcher would leave the room and the test would start.

Each label was given arbitrary numerical integer values starting from -2 for “much worse” to +2 for “much better”. Choosing a scale range of ± 2 allowed for easy comparisons between preference and perception tests. The order of playing the water sounds was randomised, but the audio-only condition was always performed before the audio-visual condition. The sounds and animations were set on auto play and after playing 6 water sounds in the audio-only condition, the presentation would stop to allow for the participant to take a short break. Then the participant needed to press Enter on the keyboard to play the next six sounds in the audio-visual condition. The test lasted between 5 and 10 minutes, including the short break.

5.6.3 Consistency test

The first 10 paired comparisons of the audio-only condition were repeated at the end of the audio-only preference test, to be used as a measure of the participant’s consistency in giving preference scores. The average consistency of all participants and the 95% confidence intervals (95% CI) were calculated. Participants who scored less than the lower bound of the 95% CI were removed from the analysis.

In this test, the average consistency score was 75% (i.e., participants gave identical results in 7.5 out of 10 repeated paired comparisons). The lower bound of the 95% CI was 70%, and hence, any participant whose consistency score was below 70% was excluded from the analysis. Only two participants had consistency scores below this limit, and their responses were excluded from the analysis of results.

5.7 Equipment and digital audio and visual processing

The audio files were all played through a pair of Beyerdynamic DT 150 closed headphones which were connected to a laptop (Acer Aspire S3 series, model No.: MS2346) through an M-Audio external USB sound card. Microsoft PowerPoint 2013 was

used to create the slides and play the audio files and the animations. The animations and the instructions were presented on a 27 in. LED monitor (Samsung LS27A350) that was connected to the laptop. The water sounds and speech signals were calibrated using a Brüel and Kjær handheld sound analyser, Type 2250. Digital audio processing was carried out using Studio One 3 audio production software (PreSonus Audio Electronics) installed on a Dell XPS L502X personal computer. Autodesk 3ds MAX with Mental Ray was used to model and render the animations. The simulation of the water particles was carried out using RealFlow 2015 (Next Limit).

5.8 Statistical analysis

The assumption of parametric tests could not be met due to the ordinal nature of the data; therefore, non-parametric tests were used to analyse the results. The data were analysed using IBM SPSS 22 for Windows. Since all participants participated in all noise conditions, Friedman's Analysis of Variance (ANOVA) was used to compare the preference and perception scores. Pairwise follow-up analysis was carried out when Friedman's ANOVA showed a significant difference among the groups or conditions. The p -values were adjusted using Benjamini-Hochberg procedure which controls the expected proportion of falsely rejected hypothesis i.e., the false discovery rate (Benjamini and Hochberg, 1995). In addition, the effect sizes of the pairwise comparisons were calculated. It is worth mentioning that both the Benjamini-Hochberg procedure and Effect size calculations are not readily given by SPSS and therefore, Microsoft Excel was used to calculate them. The program SPSS provides z -scores which can be converted to an effect size, r , using the following equation (Field, 2013):

$$r = \frac{z}{\sqrt{N}} \quad (5.2)$$

where N is the number of observations, which is twice the number of participants in this case, since the same participants took part in all test conditions.

The effects of gender, nationality, and age groups were tested using the Mann-Whitney test. Pearson's chi-square was used to perform categorical analysis on the perception ratings. Correlations between variables were tested using Spearman's correlation coefficient, r_s . The Bias-corrected and accelerated bootstrap method, BCa, was used to derive robust 95% confidence intervals, which are reported in square brackets throughout this chapter.

5.9 Results

5.9.1 Audio-only and audio-visual preference results

Normalised preference scores obtained from the paired comparisons in both audio-only and audio-visual conditions of the preference test are shown in Figure 5.4. All preference values were normalised to have values between -2 (never preferred) and +2 (always preferred).

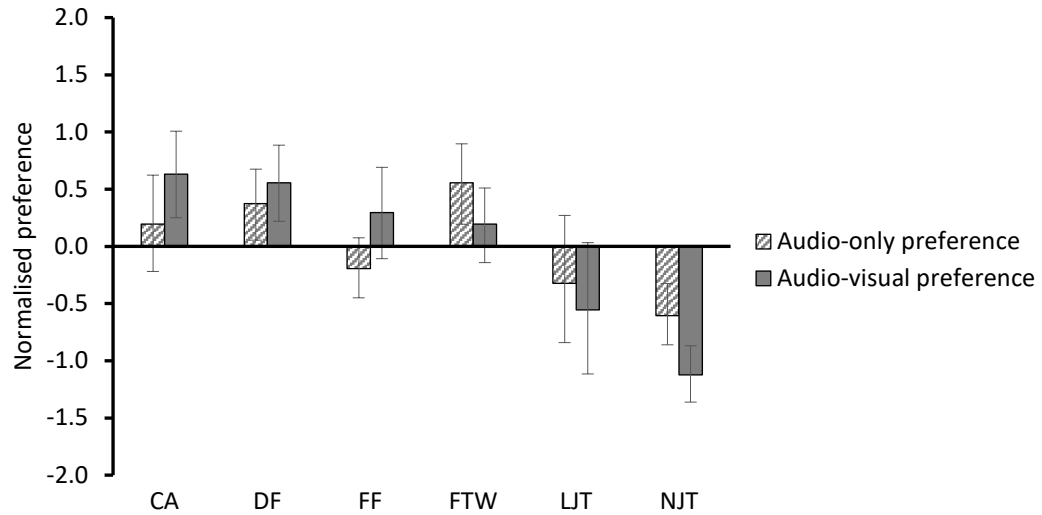


Figure 5.4 Normalised audio-only and audio-visual preference scores for six water sounds used in this study. Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals.

Audio-only preference results show the preferred water sound to be FTW ($M = 0.55$ [0.19, 0.90]), followed by DF ($M = 0.37$ [0.05, 0.68]), and CA ($M = 0.19$ [-0.22, 0.62]), respectively. The least preferred water sounds were NJT ($M = -0.61$ [-0.86, -0.33]), followed by LJT ($M = -0.32$ [-0.84, 0.27]), and FF ($M = -0.19$ [-0.45, 0.08]), respectively. Statistically, preference scores were significantly affected by the type of the water sounds, $\chi^2(5) = 18.535$, $p = .002$. Pairwise comparisons were used to follow up this finding. The results indicated that NJT was significantly less preferred than FTW ($z = -3.530$, $p < .001$, $r = -.448$), and DF ($z = -3.123$, $p = .015$, $r = -.397$). No further statistically significant differences were detected between the preference scores of the remaining water sounds.

Audio-visual preference results revealed the preferred water feature to be CA ($M = 0.63$ [0.25, 1.01]), followed by DF ($M = 0.55$ [0.22, 0.88]), FF ($M = 0.30$ [-0.11, 0.69]), and FTW ($M = 0.19$ [-0.14, 0.51]), respectively. The least preferred water feature was NJT ($M = -1.12$ [-1.36, -0.87]), followed by LJT ($M = -0.55$ [-1.12, 0.03]). Statistically, preference scores were significantly affected by the type of water feature, $\chi^2(5) = 35.472$,

$p < .001$. Pairwise comparisons were used to follow up this finding. The results indicated that NJT was significantly less preferred than CA ($z = -4.718, p < .001, r = -.599$), DF ($z = -4.480, p < .001, r = -.569$), FF ($z = -3.632, p < .001, r = -.461$), and FTW ($z = -3.724, p < .001, r = -.473$). Moreover, NJT was significantly less preferred than CA ($z = -3.157, p = .006, r = -.401$), and DF ($z = -2.919, p = .010, r = -.371$). No further statistically significant differences were detected between the preference scores of the remaining water features.

Comparing these results to those of Galbrun and Calarco (2014), it can be noted that although the rankings of the waterscapes are slightly different, the general trend of preferences is similar. In both studies, CA, DF, FF and FTW were positively rated, while NJT and LJT were negatively rated.

The addition of the visual animations seemed to help people be more confident in stating their preferences. The difference between the most and least preferred water features was larger for the audio-visual condition, in comparison to the audio-only condition. The feature that most benefited from its visual stimulus was FF which was negatively rated in the audio-only condition but was positively rated after its sound was accompanied by the visual animation. On the other hand, NJT benefited the least from its visual animation, followed by FTW. It is worth mentioning that due to the use of paired comparisons, it cannot be concluded that the visual stimuli had a negative impact on people's perception of some of the water features. The results simply suggest that some water sounds benefited more from the visual stimuli. The perception test (Section 5.9.2) further analysed whether the inclusion of the visual materials had a negative impact on people's perception of the water sounds.

In both audio-only and audio-visual conditions, no statistically significant differences were found between preference scores of males and females for all water sounds, $p > .05$.

Participants were divided into two age groups, *below 35 yr* ($N = 16$), and *35 yr and above* ($N = 15$). Preference levels towards NJT in the audio-visual condition were significantly different between the two age groups, $U = 68.00, z = -2.191, p = .041, r = -.394$. Older participants (*35 yr and above*) significantly less preferred NJT, $M = -0.800 [-1.292, -0.308]$, in comparison to younger participants $M = -1.467 [-1.740, -1.193]$. No further statistically significant differences in preference were detected between the age groups, for the remaining water sounds.

Participants were also divided into two groups based on their nationalities, *native English speakers* (N = 18) and *non-native English speakers* (N = 13). Preference levels towards FF in the audio-visual condition were significantly different between the two groups, $U = 54.00$, $z = -2.519$, $p = .011$, $r = -.452$. FF was liked by native speakers, $M = 0.756$ [0.298, 1.213], while disliked by non-native speakers, $M = -0.338$ [-1.063, -0.386]. No further statistically significant differences in preference were detected between the two groups, for the remaining water sounds.

Based on these findings, it cannot be suggested that gender, nationality and age groups of participants had significant effects on audio-only preference ratings. For the audio-visual preference, scores given to NJT was affected by the age group of participants, while scores given to FF was affected by the nationality of participants.

5.9.2 Audio-only and audio-visual perception results

The evaluation scores obtained from the audio-only and audio-visual perception tests are shown in Figure 5.5. All perception ratings were normalised to an arbitrary -2 to +2 scale with -2 representing “much worse” and +2 denoting “much better”.

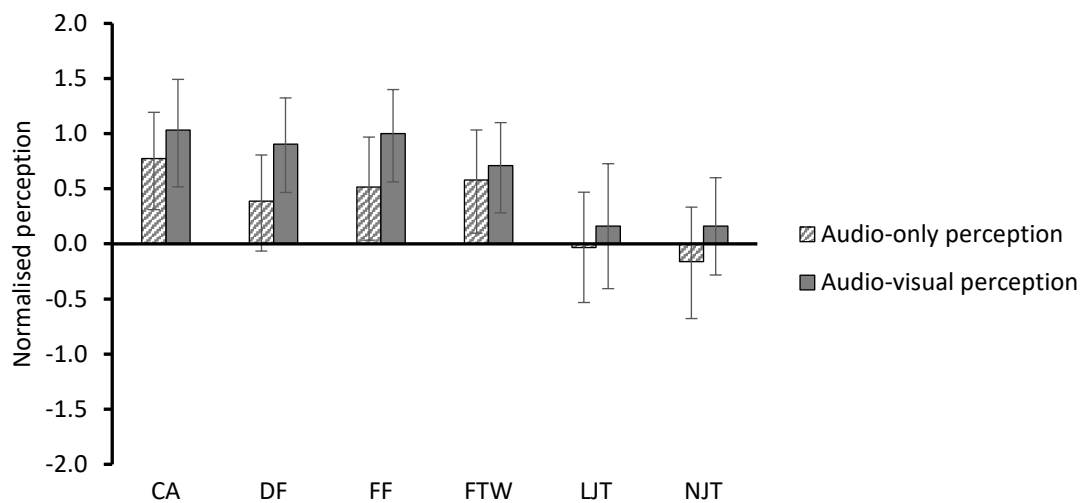


Figure 5.5 Change in participants’ perception caused by six waterscapes in audio-only and audio-visual condition. Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals.

Looking at the audio-only perception scores, it appears that four water sounds, namely, CA, DF, FF, and FTW, improved the way participants perceived the environment. The water sound that most improved participants’ perception was CA ($M = 0.77$ [0.31, 1.19]), followed by FTW ($M = 0.58$ [0.10, 1.03]), FF ($M = 0.52$ [0.03, 0.97]), and DF ($M = 0.39$ [-0.07, 0.81]), respectively. The water sound LJT ($M = -0.03$ [-0.53, 0.47]) seemed to

have had a neutral impact on the perception of people, while NJT ($M = -0.16 [-0.68, 0.33]$) had a negative impact on people's perception of their environment.

Statistically, the improvement in perception made by the water sounds was significantly different, $\chi^2(5) = 24.972$, $p < .001$. Pairwise comparisons were used to follow up this finding. Improvements made by CA ($z = 3.326$, $p = .015$, $r = .422$), FF ($z = 2.783$, $p = .038$, $r = .353$), and FTW ($z = 2.614$, $p = .038$, $r = .332$) were all significantly higher than the negative impact made by NJT. Furthermore, the improvement in perception made by CA was significantly higher than that of LJT ($z = 2.580$, $p = .038$, $r = .328$). The remaining pairwise comparisons did not reveal any statistically significant results.

In the audio-visual condition, all water features had a positive impact on the perception of participants. The water feature that was most influential was CA ($M = 1.03 [0.52, 1.49]$), closely followed by FF ($M = 1.00 [0.56, 1.40]$), then DF ($M = 0.90 [0.47, 1.32]$), FTW ($M = 0.71 [0.28, 1.10]$), LJT ($M = 0.16 [-0.41, 0.73]$), and NJT ($M = 0.16 [-0.28, 0.60]$). It is interesting that LJT and NJT, which were the least preferred water features in the preference test, still improved the way people perceived the sound environment.

Statistically, the improvement in perception made by the water features was significantly different, $\chi^2(5) = 23.602$, $p < .001$. Pairwise comparisons were used to follow up this finding. Improvements made by CA ($z = 3.021$, $p = .038$, $r = .384$), and FF ($z = 3.021$, $p = .038$, $r = .384$) were significantly higher than the change made by NJT. No further statistically significant results were detected.

In both audio-only and audio-visual conditions, no statistically significant effect of gender was detected on perception scores for all six water sounds, $p > .05$. The age of participants had a significant effect on the perception rating, only for CA, $U = 69.00$, $z = -2.123$, $p = .045$, $r = -.381$. Younger participants (*below 35 yr*) perceived CA to be significantly more beneficial, $M = 1.250 [0.794, 1.706]$ in comparison to older participants (*35 yr and above*), $M = 0.267 [-0.472, 1.006]$. No further statistically significant differences in preference were detected between the age groups, for the remaining water sounds.

Nationality (i.e., native vs. non-native English speakers) had a significant impact on the perception ratings of DF in the audio-visual condition, $U = 50.50$, $z = -2.364$, $p = .006$, $r = -.425$, and LJT in the audio-only condition, $U = 60.00$, $z = -2.807$, $p = .022$, $r = -.504$, as well as the audio-visual condition, $U = 58.00$, $z = -2.435$, $p = .018$, $r = -.437$. Native speakers, $M = 0.389 [-0.275, 1.053]$, perceived DF in the audio-only condition to be

significantly less beneficial to the sound environment than the non-native speakers, $M = 1.615$ [1.222, 2.008]. Similarly, native speakers perceived LJT to be detrimental to the sound environment in both the audio-only condition, $M = -0.556$ [-1.221, 0.110], and the audio-visual condition, $M = -0.444$ [-1.192, 0.303]. On the other hand, non-native speakers perceived LJT to positively contribute to the sound environment in both audio-only condition, $M = 0.693$ [-1.027, 1.487], and the audio-visual condition, $M = 1.000$ [0.220, 1.780]. These results suggest that perception ratings for some water sounds can be affected by whether the listener is a native English speaker, or not.

5.9.3 Comparison between audio-only and audio-visual perception scores.

As it can be seen from Figure 5.5, the average audio-visual perception scores for all water features are higher than their corresponding audio-only scores, suggesting that visual stimuli increased the level of improvement made by the water sounds alone. This was further tested for statistically significant differences between the audio-visual and audio-only scores. The results are presented in Table 5.3. The results show that the inclusion of the visual stimuli did improve the environment. This improvement was significant for two water features, namely, FF ($z = -2.950$, $p = .003$, $r = -.375$), and DF ($z = -2.311$, $p = .021$, $r = -.293$). The values of r show the magnitude of the effect that the visual stimuli had on perception. They show that FTW marginally benefited from its visual animation with a very small r value of -0.065. The effect of the animations on CA, LJT, and NJT was small with r values of -0.179, -0.125 and -0.199, respectively. On the other hand, a medium effect size was recorded for both FF, and DF with r values of -0.375, and -0.293, respectively. Using these effect sizes alongside the data presented in Table 5.3, it can be concluded that adding the visual materials further increased the improvement in perception made by the water sounds alone (audio-only condition), and the magnitude of this increase (i.e., the effect size) was “small” to “medium”, using Cohen’s (1988) scale (see Chapter 3, Section 3.5.11). This gives a rough estimate of what should be expected when a visual stimulus is added to an audio material, in terms of improving people’s perception of their environment. Attempts were made to further quantify these small to medium effect sizes, to be more interpretable. These are presented in Section 5.9.5.

Table 5.3 z-scores, p -values and effect sizes achieved by comparing audio-only and audio-visual perception scores using Wilcoxon sign-ranked test.

Sound code	CA	DF	FF	FTW	LJT	NJT
z-score	-1.407	-2.311	-2.950	-0.511	-0.981	-1.564
p -value	0.160	0.021	0.003	0.610	0.326	0.118
Effect size (r)	-0.179	-0.293	-0.375	-0.065	-0.125	-0.199

5.9.4 Comparison between preference and perception scores

The results of preference and perception tests were compared to examine whether the preferred water features were more beneficial in improving people's perception. The comparisons for both audio-only and audio-visual conditions are tabulated in Table 5.4, and presented graphically in Figure 5.6 and Figure 5.7.

For the audio-only condition (Figure 5.6), the results show that the sounds that were given positive scores in the preference test tended to improve people's perception. The only exception to this is FF which was negatively rated but positively affected people's perception. The ranking positions of the water features between the preference and perception tests are not identical. This is understandable as no significant differences in ranking could be found between the top three water features in both preference and perception tests, and therefore their rankings are likely to change in different experiments.

Table 5.4 Normalised audio-only and audio-visual scores in both preference and

Rank	Preference test				Perception test			
	Audio-only		Audio-visual		Audio-only		Audio-visual	
	Code	Pref.	Code	Pref.	Code	Perc.	Code	Perc.
1	FTW	0.55	CA	0.63	CA	0.77	CA	1.03
2	DF	0.37	DF	0.55	FTW	0.58	FF	1.00
3	CA	0.19	FF	0.30	FF	0.52	DF	0.90
4	FF	-0.19	FTW	0.19	DF	0.39	FTW	0.71
5	LJT	-0.32	LJT	-0.55	LJT	-0.03	LJT	0.16
6	NJT	-0.61	NJT	-1.12	NJT	-0.16	NJT	0.16

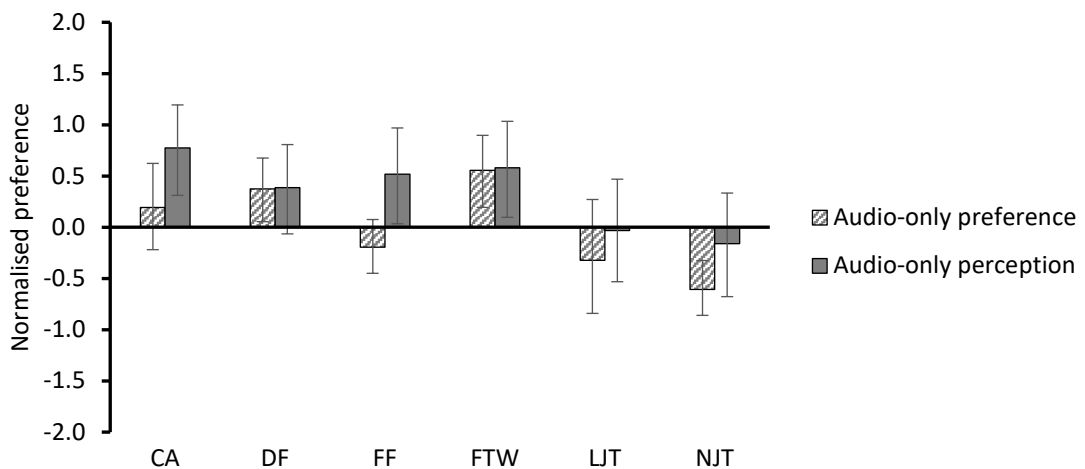


Figure 5.6 Normalised preference and perception scores given in the audio-only condition. Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals.

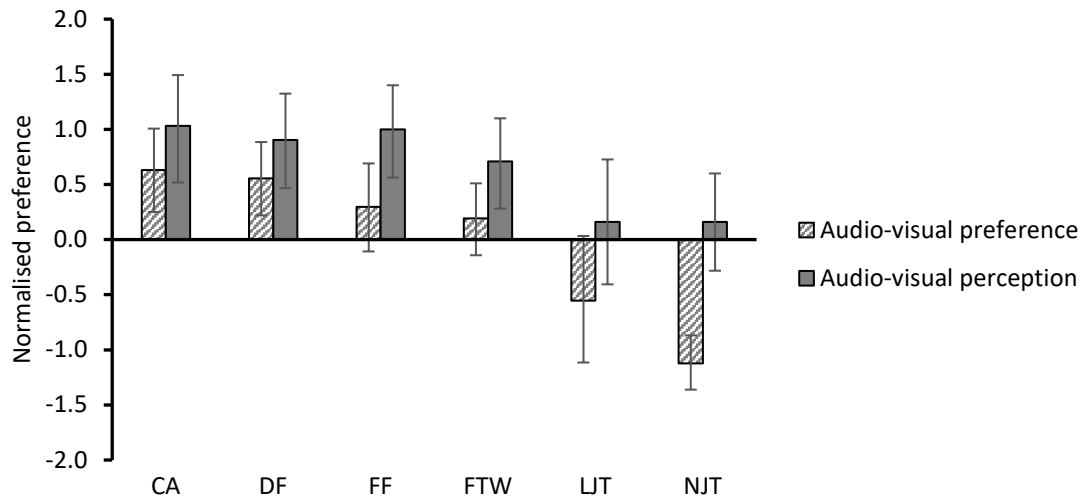


Figure 5.7 Normalised preference and perception scores given in the audio-visual condition. Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals.

For the audio-visual conditions (Figure 5.7), the water features that were positively scored in the preference test (i.e., CA, DF, FF, and FTW), were highly rated in the perception test, too. The least preferred water features (i.e., LJT and NJT) improved people's perception the least, yet still positively contributed to changing the sound environment.

Table 5.5 Spearman's correlation (r_s) between preference and perception scores for each waterscape in both audio-only and audio-visual conditions.

Code	Audio-only tests						Audio-visual tests					
	CA	DF	FF	FTW	LJT	NJT	CA	DF	FF	FTW	LJT	NJT
r_s	-.271	.093	.083	.003	.657**	.349	.198	.130	-.008	.165	.542*	.390*
p -value	.141	.617	.656	.988	.000	.054	.285	.485	.967	.375	.002	.030
95% CI												
Low	-.600	-.280	-.267	-.373	.420	-.038	-.187	-.287	-.335	-.184	.247	.028
High	.080	.425	.444	.390	.823	.664	.554	.483	.316	.511	.776	.681

* Correlation is significant at the 0.05 level. ** Correlation is significant at the 0.001 level. CI Low = 95% BAc lower bound confidence intervals. CI High = 95% BAc higher bound confidence intervals.

Correlations between preference and perception test scores were tested using Spearman's correlation coefficient (r_s) for both audio-only and audio-visual conditions. The results, reported in Table 5.5 (above), indicate that there was a positive correlation between preference and perception scores for all water sounds in the audio-only condition, except CA, for which the correlation was negative. Similarly, in the audio-visual condition, there was a positive correlation between preference and perception scores for all water sounds except for FF, for which the correlation was negative. However, the magnitude of the correlations was small in most cases and significant values could only be achieved for

LJT in both audio-only, $r_s = .657$, $p < .001$, and audio-visual, $r_s = .542$, $p = .002$, conditions, and for NJT in the audio-visual condition, $r_s = .390$, $p = .030$.

These results suggest that, generally, there is a reasonable agreement between perception and preference scores, despite the variation in rankings of the waterscapes, indicating that the water features that tend to be preferred were likely to be beneficial in view of improving people's perception of their environment. However, using both tests results in more conclusive and precise findings, as the two tests do not necessarily measure the same construct. This was shown by the weak and mainly non-significant correlations between the two tests.

5.9.5 *Categorical analysis of perception scores*

The previous analysis of the data using Friedman's ANOVA and Wilcoxon test identified the differences among perception scores of the water features. To further examine how meaningful and practical these differences were, a categorical analysis was performed using the Chi-square test. For a water feature to be accepted as being practically beneficial in improving people's perception, the number of people who positively perceived the water sounds must be significantly higher than those who perceived it as being detrimental.

In order to allow for this analysis to be performed, for each water sound in each of the audio-only condition and the audio-visual condition, the frequency of selecting labels "slightly better" and "much better" in the perception test was calculated and categorised as *positive scores*. Similarly, the frequency of selecting labels "slightly worse" and "much worse", was calculated and categorised as *negative scores*. Then, the frequencies of the two categories (i.e., positive scores and negative scores) were compared using the Chi-square test which allows for this type of analysis to be performed. Any statistically significant difference between the two groups would mean the magnitude of improvement or detriment in people's perception is significant, and therefore meaningful. The results of the Chi square test are shown in Table 5.6.

In the audio-only condition, for CA, DF, FF and FTW (i.e., positive average perception scores; refer to Figure 5.5) any significant result would mean that the water sound meaningfully improved the environment. On the contrary, a significant result for LJT and NJT, which were negatively perceived, would mean that the water sound significantly deteriorated the environment. In the audio-visual condition, any significant score would

indicate that the water feature significantly improved people's perception, as the average perception scores for all water features were positive.

In the audio-only condition, two water sounds significantly improved the environment. The water sounds were CA, $\chi^2(1) = 7.759, p = .008$, and FTW, $\chi^2(1) = 5.452, p = .029$. Two more water sounds, namely DF and FF, improved the environment but the level of improvement was not statistically significant, $p > .05$. Water sounds LJT and NJT deteriorated the environment (as their mean perception scores were negative), yet no significant results could be detected, $p > .05$. These results suggest that two water sounds significantly improved the environment, and none significantly deteriorated the environment, even if negatively perceived.

In the audio-visual condition, four water features resulted in a significant improvement in people's perception. The four water features were CA, $\chi^2(1) = 14.286, p < .001$, FF, $\chi^2(1) = 12.448, p = .001$, DF, $\chi^2(1) = 10.704, p = .002$, and FTW, $\chi^2(1) = 9.846, p = .002$. No significant results could be recorded for LJT and NJT, $p > .05$, which indicate that despite being positively perceived, the level of change in perception was not statistically significant, and therefore, practically meaningless.

It is worth noting that the visual stimuli helped to increase the tendency to perceive the environment in a positive way for all water sounds. To further quantify this, the odds ratio between audio-only and audio-visual scores was calculated. The odds ratio is an effect size which quantifies the relationship between variables (Field, 2013). For each water sound, the number of positive scores was divided by the number of negative scores, in both audio-only and audio-visual conditions. This resulted in two ratios. Then the ratio from the audio-visual condition was divided by the ratio from the audio-only condition, to result in the odds ratio for that water sound. Any odds ratio larger than 1 would suggest that the visual materials increased the likelihood of obtaining positive scores. On the contrary, odds ratios smaller than 1 would suggest that the visual materials increased the likelihood of obtaining negative scores. The odds ratio for each water sound is provided at the bottom of Table 5.6. All odds ratios are larger than 1, which confirms the previous finding that visual stimuli did have a positive impact on people's perception of their environment. The odds ratio can also be used to quantify the "small" to "medium" effect sizes reported in Section 5.9.3. For instance, CA has an odds ratio of 1.9, which suggests that the addition of the visual material to this sound increased the probability of getting positive changes in perception by 1.9 times. The highest odds ratio was recorded for FF,

which was 2.5, suggesting that adding a visual stimulus to this water sound increased the chance of obtaining positive scores by approximately 2.5 times. The lowest odds ratio was recorded for LJT which was 1.1. Hence, within the context of the perception test, it can be suggested that a small to medium effect size would mean approximately 1.1 to 2.5 times increase in the probability of making positive changes in the way people perceive their environment.

Table 5.6 Chi-square (χ^2) test statistic and odds ratios between audio-only and audio-visual scores in the perception test.

	Audio-only condition					
Sound code	CA	DF	FF	FTW	LJT	NJT
Chi-Square (χ^2)	7.759*	4.481	2.793	5.452*	0.000	0.143
df	1	1	1	1	1	1
p-value	.008	.052	.136	.029	1.000	.851
	Audio-visual condition					
Sound code	CA	DF	FF	FTW	LJT	NJT
Chi-Square (χ^2)	14.286**	10.704**	12.448**	9.846**	0.133	0.333
df	1	1	1	1	1	1
p-value	.000	.002	.001	.002	.856	.701
Odds ratio	1.909	1.853	2.526	1.718	1.143	1.442

df = degree of freedom. * χ^2 is significant at the 0.05 level. ** χ^2 is significant at the 0.001 level.

5.9.6 Cluster analysis

The results obtained from the preference test were further analysed using Kendall's coefficient of concordance, W , to look for agreement among participants. Kendall's W results in a score between 0 (no agreement) to 1 (perfect agreement). The level of agreement for the audio-only preference test was $W = .120$, $p < .001$. For the audio-visual preference scores, the agreement among participants was slightly higher, $W = .229$, $p < .001$. In both conditions, the levels of agreement were very low (i.e., values closer to 0 than 1) which suggests having clusters of participants who rated the water sounds differently. These low levels of agreement among participants led to further analysis as participants might have come from different groups and did not necessarily share the same preferences.

Hierarchical cluster analysis was adopted in view of identifying more consistent clusters of participants. *Ward's method* was applied as a clustering method and the *Square Euclidian distance* was used as the interval measure. The results are tabulated in Table 5.7 and they show that participants can be split into two clusters according to their preference scores, with both clusters showing different values of Kendall's W . The audio-

only and the audio-visual preference scores in Cluster 1 and Cluster 2 are shown in Figure 5.8 and Figure 5.9, respectively.

Table 5.7 Audio-only and audio-visual preference scores and rankings of six waterscapes used in this study, from all participants retained for the analysis and from clusters obtained from the hierarchical cluster analysis. The preferences are listed as normalised preference values. Kendall's coefficient of concordance, W , is also given for results including all subjects and for the clusters, as well as age and gender distributions of participants in each cluster.

Rank	All subjects				Cluster 1 (17 participants)				Cluster 2 (14 participants)			
	Audio-only		Audio-visual		Audio-only		Audio-visual		Audio-only		Audio-visual	
	code	Pref.	code	Pref.	Code	Pref.	Code	Pref.	Code	Pref.	Code	Nor.
1	FTW	0.55	CA	0.63	CA	0.78	CA	1.20	LJT	1.37	LJT	0.80
2	DF	0.37	DF	0.55	FTW	0.96	DF	0.87	FTW	0.06	FTW	0.23
3	CA	0.19	FF	0.30	DF	0.82	FF	0.78	DF	-0.17	DF	0.17
4	FF	-0.19	FTW	0.19	FF	-0.12	FTW	0.16	FF	-0.29	CA	-0.06
5	LJT	-0.32	LJT	-0.55	NJT	-0.73	NJT	-1.34	NJT	-0.46	FF	-0.29
6	NJT	-0.61	NJT	-1.12	LJT	-1.72	LJT	-1.67	CA	-0.51	NJT	-0.86
W	0.120		0.229		0.639		0.706		0.329		0.153	
Number of participants					17				14			
Male					9				7			
Female					8				7			
Native					12				7			
Non-native					5				9			
Mean age [95% CI] (yr)					35.7 [31.7, 39.8]				37.1 [30.5, 43.6]			

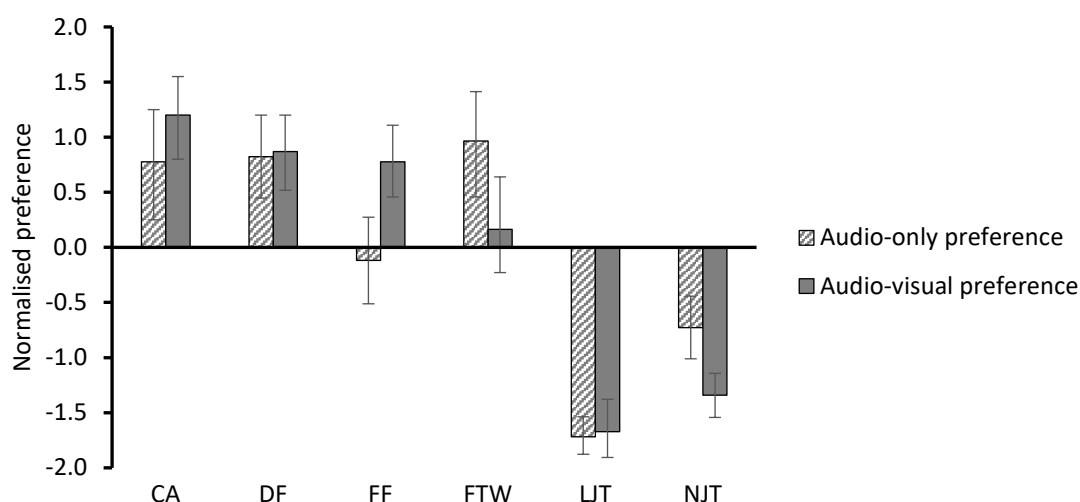


Figure 5.8 Normalised audio-only and audio-visual preference scores in Cluster 1 (17 participants). Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals.

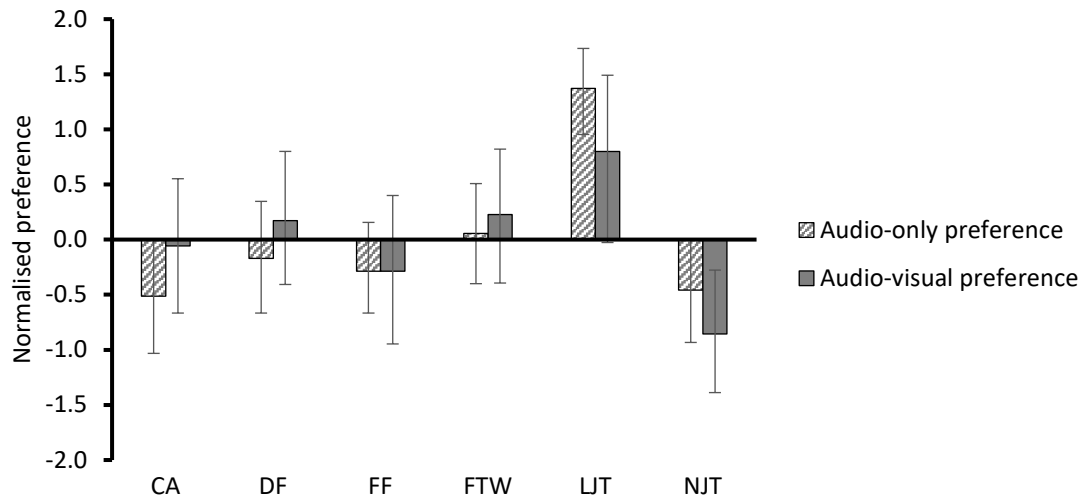


Figure 5.9 Normalised audio-only and audio-visual preference scores in Cluster 2 (14 participants). Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals.

For the audio-only preference, the agreement among participants increased for both Cluster 1, $W = .639$, $p < .001$, and Cluster 2, $W = .329$, $p < .001$, in comparison to the original level of agreement, $W = .120$, $p < .001$, when participants were not clustered. The water features that were most affected by the clustering were LJT and CA with a variation of ± 5 in ranking positions between the two clusters. In Cluster 1, the most preferred water sound was CA, and the least preferred water sound was LJT. In cluster 2, the exact opposite was observed, as LJT became the most preferred water sound and CA the least preferred water sound.

Results obtained for the audio-visual preference clustering showed that participants grouped under Cluster 1 had a high level of agreement, $W = .706$, $p < .001$, while participants under Cluster 2 had a much lower level of agreement $W = .153$, $p = .057$. The water feature most affected by clustering was LJT, with ± 5 changes in ranking positions, followed by CA, with ± 3 variations in ranking positions between the two clusters. The most preferred water feature in Cluster 1 was CA, and the least preferred water feature was LJT. In cluster 2, the most and the least preferred water features were LJT and NJT, respectively. The preference scores given to LJT in both audio-only and audio-visual conditions are interesting. It appears that LJT was either liked or disliked and nowhere in between.

It can also be noticed that the variation in preference levels in Cluster 1 is distributed over a wider range. For instance, the most liked water feature in the audio-visual test was CA

with a normalised preference value of 1.20, whilst the least preferred water feature was LJT with a normalised preference score of -1.67 (i.e., a range of 2.87). This range was smaller for Cluster 2, with LJT, the most preferred water feature having a normalised preference score of 0.80, and NJT, the least preferred, having a score of -0.86 (i.e., a range of 1.66). In addition, the distribution of preference scores in Cluster 2 given to CA, DF, FF, and FTW looks random with values closer to zero than ± 2 (mainly smaller than ± 0.50 except for CA in the audio-only condition with a value of -0.51). These suggest that people in Cluster 2 might have struggled to state their preferences, and therefore might have changed their preferred soundscape throughout the test. Another explanation is that participants might have equally preferred most water sounds, but because of using paired comparisons, they were required to choose an option, and therefore, they might have randomly selected one. Apart from LJT which was distinctly more preferred, it would appear that participants in Cluster 2 equally preferred the remaining waterscapes, and hence the randomness in their preference scores. This became more apparent in the perception test.

To further analyse the above assumption, the audio-only and audio-visual preference scores for participants in Cluster 2 were tested for statistical differences using Friedman's ANOVA test. The audio-only preference scores were significantly different, $\chi^2(5) = 23.000, p < .001$. Pairwise comparisons were used to follow up this finding, which revealed that LJT was significantly more preferred than all remaining water sounds, CA ($z = -3.889, p < .001, r = -.735$), NJT ($z = 3.738, p < .001, r = .706$), FF ($z = -3.536, p < .001, r = -.668$), DF ($z = -3.232, p < .001, r = -.611$), and FTW ($z = -2.576, p < .001, r = -.487$). However, pairwise comparisons between the remaining water sounds were all non-significant, $p > .05$. The audio-visual preference scores in Cluster 2 did not reveal a significant difference, $\chi^2(5) = 10.737, p = .057$, i.e., water features received similar preference scores. These findings indicate that people in Cluster 2 equally preferred CA, DF, FF, FTW, and NJT.

The same clustering criteria used in the preference test were used to cluster the perception scores into two clusters. Perception scores and rankings of the water features, in both audio-only and audio-visual conditions, are presented in Table 5.8. In both audio-only and audio-visual conditions, participants who were in Cluster 1 in the preference test, perceived four water sounds, namely, CA, DF, FF, and FTW to positively affect their perception of the environment, while LJT and NJT were perceived as having an adverse effect on people's perception. This result is presented in Figure 5.10.

Participants in Cluster 2 perceived all water sounds to be beneficial, in both audio-only and audio-visual conditions. This result is shown in Figure 5.11. The ratings of the water sounds in Cluster 2 are comparable with the exception of LJT which received a distinctly higher perception score. This trend explains the randomness in preference scores in this cluster and confirms that people in this cluster perceived all water sounds to be beneficial, with LJT receiving the highest score.

Table 5.8 Audio-only and audio-visual perception scores and rankings of six waterscapes used in this study, from all participants retained for the analysis and from clusters obtained from the hierarchical cluster analysis. Scores are listed as normalised perception values. Age and gender distributions of participants in each cluster are also provided.

Rank	All subjects				Cluster 1				Cluster 2			
	Audio-only		Audio-visual		Audio-only		Audio-visual		Audio-only		Audio-visual	
	code	Pref.	code	Pref.	Code	Pref.	Code	Pref.	Code	Pref.	Code	Nor.
1	CA	0.77	CA	1.03	CA	0.76	CA	1.24	LJT	1.07	LJT	1.36
2	FTW	0.58	FF	1.00	FF	0.53	FF	1.00	CA	0.79	FF	1.00
3	FF	0.52	DF	0.90	FTW	0.41	DF	0.88	FTW	0.79	DF	0.93
4	DF	0.39	FTW	0.71	DF	0.29	FTW	0.71	DF	0.50	CA	0.79
5	LJT	-0.03	LJT	0.16	NJT	-0.71	NJT	-0.29	FF	0.50	FTW	0.71
6	NJT	-0.16	NJT	0.16	LJT	-0.94	LJT	-0.82	NJT	0.50	NJT	0.71
Number of participants					17				14			
Male					9				7			
Female					8				7			
Native					12				7			
Non-native					5				9			
Mean age [95% CI] (yr)					35.7 [31.7, 39.8]				37.1 [30.5, 43.6]			

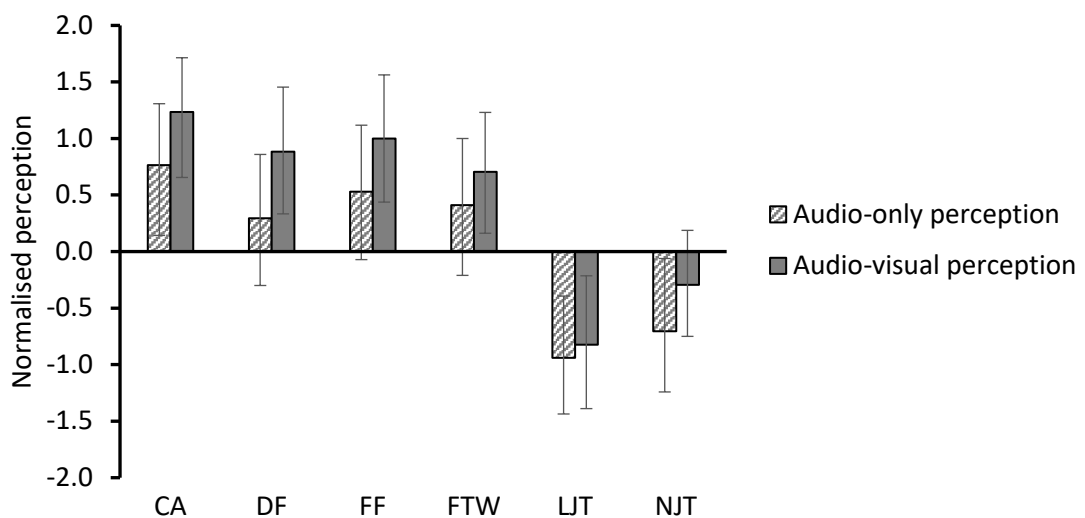


Figure 5.10 Normalised audio-only and audio-visual perception scores in Cluster 1 (17 participants). Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals.

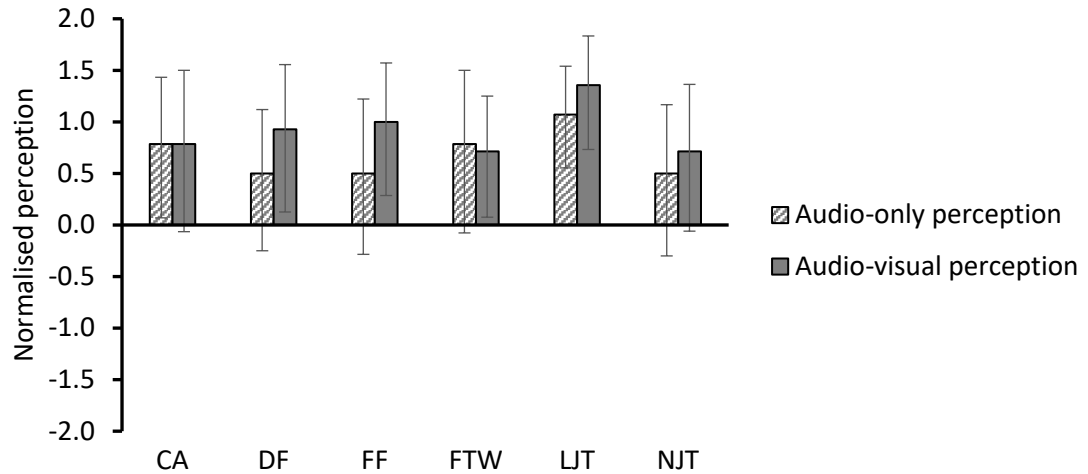


Figure 5.11 Normalised audio-only and audio-visual perception scores in Cluster 2 (14 participants). Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals.

5.9.7 Correlation between preference/perception ratings and acoustic/psychoacoustic measures

The audio-only preference and perception ratings were further analysed to examine if the ratings could be attributed to an objective measure, instead of semantic descriptors. The correlation analysis allowed for examining the relationship between the preference/perception ratings and some measures that have been reported in previous research to correlate well with preference and tranquillity ratings of water generated sounds. Four acoustic measures and three psychoacoustic measures were examined. The acoustic measures were L_{10} - L_{90} , L_{Ceq} - L_{Aeq} , the Low-High Frequency A-weighted Level Difference (L_{AeqLow} - $L_{AeqHigh}$), and the sound interference level (SIL). The psychoacoustic measures were sharpness, roughness, and pitch strength. L_{10} - L_{90} represents temporal variations in a sound, and has been reported to have a positive correlation with preference ratings of water sounds (Watts *et al.*, 2009; Galbrun and Ali, 2013; Galbrun and Calarco, 2014). L_{Ceq} - L_{Aeq} shows how dominant the lower frequencies are in a sound, and has been reported to positively correlate with preference ratings of water sounds used to mask road traffic noise (Galbrun and Ali, 2013). The sound interference level (SIL) is the arithmetical mean of the sound pressure levels at 500, 1k and 2k Hz, and was suggested by Watts *et al.* (2009) to have a positive correlation with tranquillity levels of water generated sounds. The Low-High Frequency A-weighted Level Difference (L_{AeqLow} - $L_{AeqHigh}$) was proposed by Veitch *et al.* (2002) which is the difference between the A-weighted sound pressure level of lower frequencies (16-500 Hz), and the higher frequencies (1k-16k Hz). This measure has been shown to correlate well with acoustic

satisfaction in open-plan offices, when an artificially generated masking sound (e.g., a filtered pink noise) is used as a speech masking sound (Veitch *et al.*, 2002; Hongisto *et al.*, 2015). Among the psychoacoustic measures, sharpness has been reported to closely correlate with changes in tranquillity levels of water sounds (Watts *et al.*, 2009), as well as their preference ratings (Jeon *et al.*, 2012; Galbrun and Ali, 2013). Pitch strength and roughness were both correlated well with semantic components of water sounds in Galbrun and Calarco's (2014) study.

The results of the correlation analysis of the preference test are shown in Table 5.9. When scores from all participants were taken into consideration, sharpness, and pitch strength positively correlated with the preference ratings, while the remaining measures had a negative correlation with the preference ratings. However, the correlation was statistically significant only for roughness, $r_s = -.870$, $p = .024$.

Previous research (e.g., Galbrun and Calarco, 2014) has shown that preferred water sounds are those that are less likely to be confused with man-made sounds such as toilets and sinks. These preferred water sounds tend to have low levels of sharpness and pitch strength. Therefore, the positive correlation between sharpness and preference and between pitch strength and preference might be associated with the preferred water sounds being unambiguously perceived as being natural.

The same pattern of correlations was observed when the analysis was repeated based on preference scores from participants in Cluster 1. Sharpness and pitch strength maintained a positive correlation and the remaining measures had a negative correlation with preference ratings, and the only statistically significant result was for roughness, $r_s = .928$, $p = .008$.

When the analysis was performed on Cluster 2, the pattern of correlations changed. Sharpness, which had a positive correlation in Cluster 1, showed a negative correlation with preference ratings. Whilst, $L_{10}-L_{90}$, $L_{Ceq}-L_{Aeq}$, SIL and $L_{AeqLow}-L_{AeqHigh}$, all showed a positive correlation with preference ratings, despite having a negative correlation in Cluster 1. The only two measures for which the direction of correlation did not change between the two clusters were roughness (with a negative correlation) and pitch strength (with a positive correlation).

The analysis shows how unreliable these measures can be in predicting preference ratings of water sounds, as the direction of correlation (positive and negative) tended to change between the clusters. The only two measures that seemed to be more reliable than others

were roughness, and pitch strength, mainly because their direction of correlation remained the same irrespective of clusters. Since only six water sounds were tested in this study (a small sample size), whether the correlations are significant or not, is not as important as the direction of the correlation between the clusters. Based on these results, it can be suggested that water sounds with low values of roughness and high values of pitch strength tend to be preferred.

Table 5.9 Spearman's correlation (r_s) between participants' preference scores in the audio-only condition and some acoustic and psychoacoustic measures. r_s values are shown in bold.

Measures			$L_{10}-L_{90}$	$L_{Ceq}-L_{Aeq}$	Sharpness [†]	Roughness ^{††}	Pitch strength	SIL ^{†††}	$L_{AeqLow}-L_{AeqHigh}$
All participants	r_s		-.657	-.029	.493	-.870*	.464	-.371	-.429
	p -value		.156	.957	.321	.024	.354	.468	.397
	BCa	Lower limit	-1	-.935	-.948	-1	-.894	-1	-1
	95% CI	Upper limit	1	.730	1	-.168	1	.806	.500
Cluster 1	r_s		-.771	-.257	.667	-.928**	.406	-.486	-.600
	p -value		.072	.623	.148	.008	.425	.329	.208
	BCa	Lower limit	-1	-1	-.333	-1	-1	-1	-
	95% CI	Upper limit	-.091	.806	1	-.515	1	.500	.091
Cluster 2	r_s		.486	.829*	-.493	-.058	.667	.429	.600
	p -value		.329	.042	.321	.913	.148	.397	.208
	BCa	Lower limit	-.818	.091	-1	-1	0	-.636	-.515
	95% CI	Upper limit	1	1	.500	.800	.894	1	1

[†]Sharpness is in acum. ^{††}Roughness is in asper. ^{†††}Sound interference level. *Correlation is significant at the 0.05 level. **Correlation is significant at the 0.01 level.

Table 5.10 Spearman's correlation (r_s) between participants' perception scores in the audio-only condition and some acoustic and psychoacoustic measures. r_s values are shown in bold.

Measures			$L_{10}-L_{90}$	$L_{Ceq}-L_{Aeq}$	Sharpness [†]	Roughness ^{††}	Pitch strength	SIL ^{†††}	$L_{AeqLow}-L_{AeqHigh}$
All participants	r_s		-.714	-.429	.580	-.493	-.261	-.486	-.600
	p -value		.111	.397	.228	.321	.618	.329	.208
	BCa	Lower limit	-1	-1	-.800	-.894	-.949	-1	-1
	95% CI	Upper limit	.598	.600	1	.678	1	1	.869
Cluster 1	r_s		-.657	-.600	.551	-.464	-.522	-.371	-.600
	p -value		.156	.208	.257	.354	.288	.468	.208
	BCa	Lower limit	-1	-1	-.394	-1	-1	-1	-1
	95% CI	Upper limit	.757	.806	1	.978	.548	1	.673
Cluster 2	r_s		0	0.154	0	.391	.047	-.247	0
	p -value		1	.770	1	.443	.930	.637	1
	BCa	Lower limit	-.894	-.905	-.787	-.775	-.887	-.905	-.894
	95% CI	Upper limit	.787	.910	.949	1	.826	.572	.787

[†]Sharpness is in acum. ^{††}Roughness is in asper. ^{†††}Sound interference level. *Correlation is significant at the 0.05 level. **Correlation is significant at the 0.01 level.

For the perception test, as shown in Table 5.10, no statistically significant correlation was found between the perception ratings and the acoustic and psychoacoustic measures. The direction of the correlation tended to change between the two clusters for all measures except the SIL, which maintained a negative correlation with perception ratings, regardless of the clusters.

5.10 Discussion

Six water features were tested in this experiment under two different conditions. The general trend showed 4 water features to be preferred and to improve people's perception of their environment. The four water features were CA, DF, FF and FTW. The most preferred water sound in the audio-only condition was FTW, while the most preferred water feature in the audio-visual condition was CA. Comparing these findings to those reported by Galbrun and Calarco (2014), there are more similarities than differences between the two studies, despite using two different background noises (i.e., road traffic noise vs. irrelevant speech) to be masked by the water sounds. In both studies, CA, DF, FF and FTW were highly preferred, while NJT was poorly rated. Both studies found that people can have opposing preferences to LJT, which tends to be either very highly rated or very poorly rated. Furthermore, it appears that people's preferences were not greatly affected by the background noise which the water sounds masked. Galbrun and Calarco (2014) used road traffic noise whilst speech was used in the current study, yet the preference results are comparable. Hence, it is likely that the findings of this study are applicable to other background noises such as those found in hotel lobbies and supermarkets, for example.

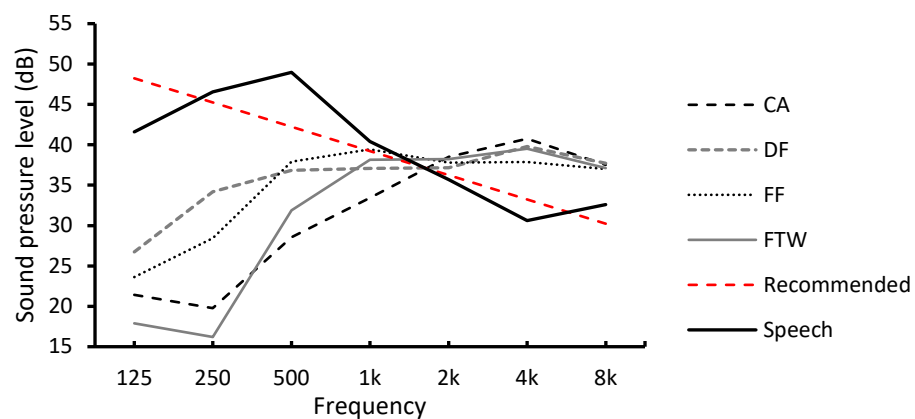


Figure 5.12 Octave-band spectra of speech (48 dBA) and four highly preferred water sounds in the preference test, against the reference masking spectrum recommended by Veitch *et al.* (2002), calibrated to 45 dBA.

The spectra of the four preferred water sounds alongside the recommended masking spectrum are shown in Figure 5.12. The widely accepted spectrum of a masking sound is a pink noise whose sound pressure level reduces by 5 dB per octave band (Veitch *et al.*, 2002). As Figure 5.12 shows, spectra of the water sounds do not resemble the recommended spectrum (Veitch *et al.*, 2002), yet people still rated them positively. The recommended spectrum was based on using artificial masking sounds, and therefore, might not be applicable to natural sounds.

Comparing the octave-band spectra of these water sounds to that of speech reveals that the sound energy of the water sounds was more concentrated at the middle to high-frequency range, while the speech contained more sound energy at a lower frequency range. For instance, at 500 Hz, the sound pressure level of speech is 11 dB higher than that of the loudest water sound at that frequency (i.e., FF). This suggests that the water sounds were not capable of masking speech at frequencies below 1 kHz. At higher frequency bands (2 kHz and 4 kHz) the water sounds have more energy than the speech signal which suggests that speech masking might have happened at those frequency bands. However, the general trend implies that there might have been more informational masking than energetic masking. The water sounds might have masked selective phonemes instead of entire words, especially consonant sounds which are characterised by having high-frequency contents.

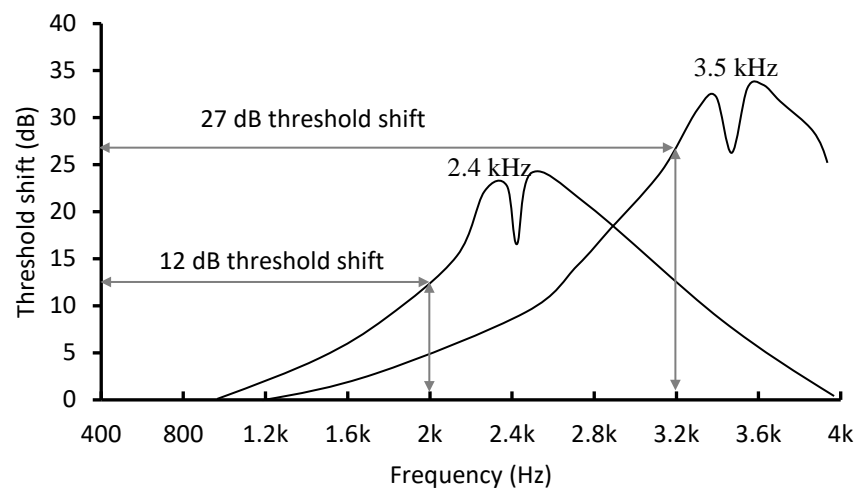


Figure 5.13 Masking curves at 40 dB for two pure tones, 2.4 kHz and 3.5 kHz (Fletcher, 1953) (Fig. reproduced).

Going back to the fundamentals of sound masking, a masking sound changes the threshold of audibility of the masked sounds (Long, 2014). For instance, in the case of

masking a pure tone by another pure tone, for the masked tone to become audible, it would need to be played at a higher level (i.e., higher than its normal threshold of audibility) in the presence of the masker tone. Early research carried out in the 1950s determined the masking curves for pure tones at various frequencies and sound levels (Fletcher, 1953). Due to their relevance to the current study, the masking curves of a pure tone at 40 dB and at two frequencies, 2.4 kHz and 3.5 kHz, are represented in Figure 5.13.

The curves show the difference between the normal threshold of audibility of a sound and its new threshold in the presence of the masking tone. For example, for the 2.4 kHz masking tone, at 40 dB, it induces a 12 dB threshold shift in a 2 kHz tone. For the 3.5 kHz masking tone, it induces a 27 dB threshold shift in a 3.2 kHz tone. The preferred water sounds have sound pressure levels close to 40 dB between 2 kHz and 8 kHz octave bands, which suggests that they might have been capable of shifting the threshold of audibility of speech at those frequency bands by up to 35 dB.

Human speech contains vowel and consonant sounds. The consonant sounds are characterised by having more high-frequency contents. The spectrograms of two types of consonants, plosives and fricatives are shown in Figure 5.14, and Figure 5.15, respectively. Plosives (Figure 5.14) are consonants that are produced by stopping the air flow using the articulators (i.e., lips, teeth, and palate) followed by a sudden release of air. Examples of plosive consonants are /d/ and /t/. Fricatives (Figure 5.15) are consonants that are made by restricting the flow of air to pass through a narrow channel made for example by the two lips so that a friction-sound is produced. Examples of fricative consonants are /f/, /s/ and /z/. As shown in Figure 5.14, and Figure 5.15, these consonants have high-frequency contents (shown by darker shades and marked by a rectangle) between 3 kHz and 4 kHz, where the water sounds have their highest sound pressure levels (around 40 dB). At that sound level, the water sounds can change the threshold of audibility of these consonants by up to 35 dB. Hence, it is likely that the water sounds masked selected sounds in the speech signal especially the fricative and the plosives consonants, instead of masking the entire word or sentence, which could have changed the meaning of the speech and made it more difficult to understand. This explanation overly simplifies the speech signal as speech is much more complex than a pure tone, but it provides an insight into understanding the informational masking caused by the water sounds.

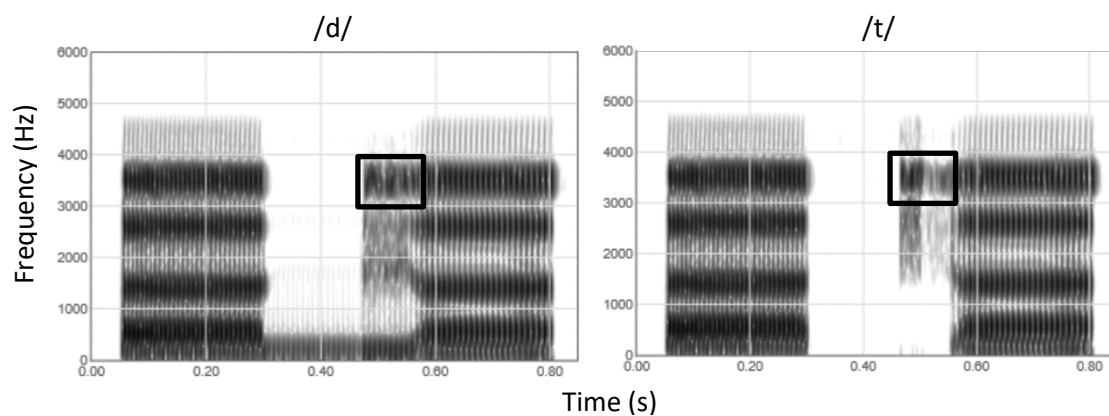


Figure 5.14 Spectrogram of two plosives consonants spoken between two vowel sounds. High-frequency contents (dark shades) are marked by a black rectangle (UCL, 2017).

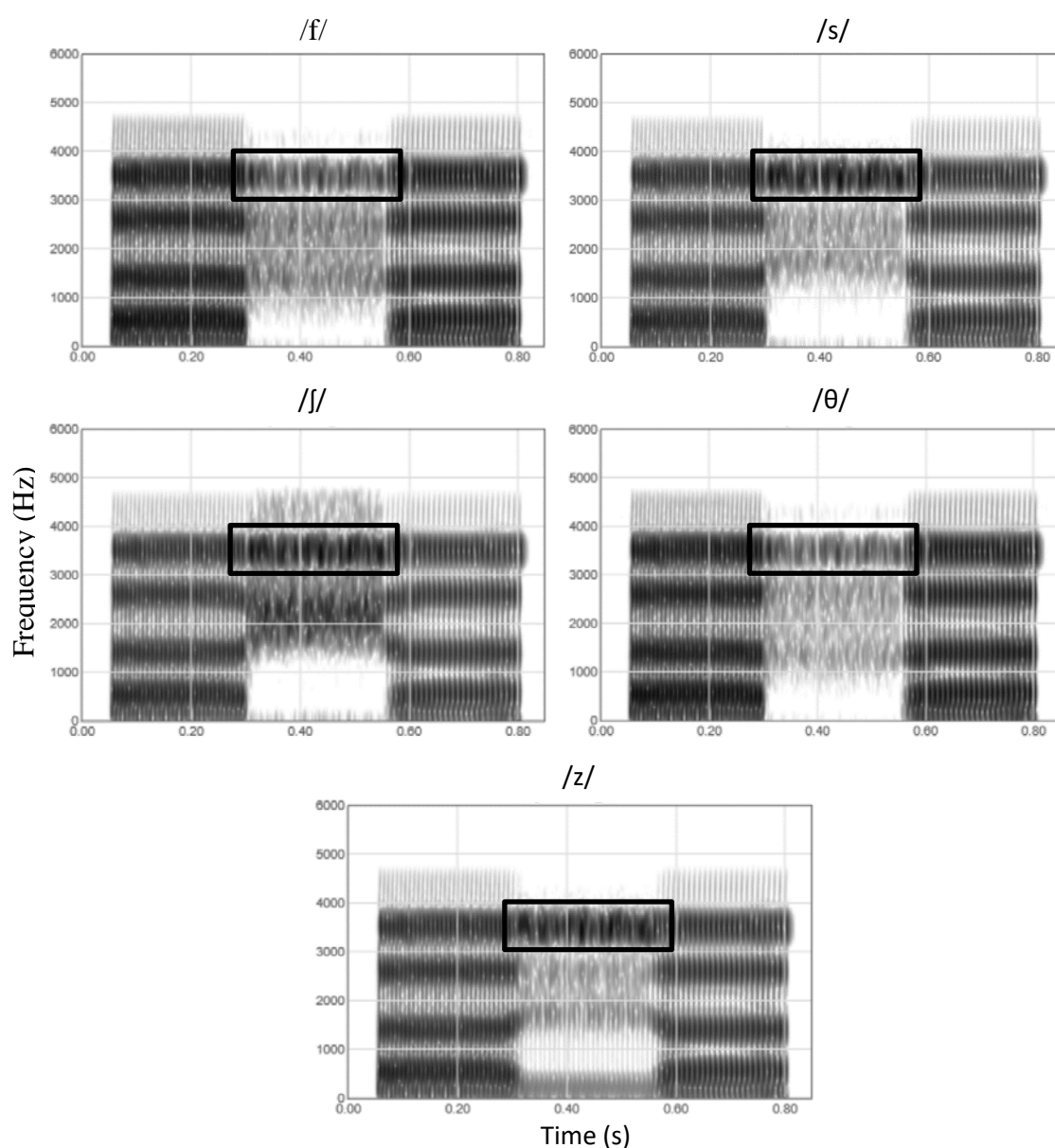


Figure 5.15 Spectrogram of five fricative consonant spoken between two vowel sounds. High-frequency contents (dark shades) are marked by a black rectangle (UCL, 2017).

The addition of the visual materials seemed to have helped people to be more confident in stating their preferences, and also increased the chance of obtaining positive changes in people's perception by up to 2.5 times. The visual materials never had a detrimental effect on people's perception. Given the quality and lifelikeness of the visual animations used in this study, the credibility of findings obtained in the audio-visual preference and perception tests is likely to be higher than previously published studies, in which either still images of water features or video recordings were used. A still image cannot show the movement and liveliness of water, and a video recording is less controllable when examining different water features, as the background of the video will typically change. Therefore, it is safe to state that the findings of this study support using a real water feature over just playing its sound through speakers.

The correlation analysis revealed that only roughness had a statistically significant correlation with preference ratings of the water sounds, and no measure significantly correlated with the perception ratings of the water sounds. When the analysis was repeated based on clusters, the measures tend to revert their direction of correlation between the clusters, which suggests the unreliability of the measure used in this study in predicting preference and perception ratings of water sounds. The only two measures which maintained their direction of correlation with the preference ratings, between the clusters, were roughness, with a negative correlation, and pitch strength, with a positive correlation. Furthermore, despite small correlations, the sound interference level (SIL) was the only measure which maintained a negative correlation with perception ratings in both clusters. All the remaining measures had positive correlations in one cluster and negative correlations in the other. Hence, future waterscape studies should pay more attention to roughness, pitch strength and to some extent, the sound interference level.

5.11 Conclusions

Six water sounds were used to mask irrelevant speech. These water sounds were a 4-step cascade (CA), a dome fountain (DF), a foam fountain (FF), a 37-jet fountain (FTW), a large jet (LJT) and a narrow jet (NJT). Two tests were carried out, a preference test and a perception test. Both tests included audio-only (no animations) and audio-visual (with animations) conditions. The audio-only preference scores revealed FTW as the most preferred water sound and NJT as the least preferred water sound. In the audio-visual preference scores, CA was the most preferred water feature and the least preferred water feature was NJT. Water sounds benefited differently from their visual stimulus. The

feature that most benefited from its visual stimulus was FF whilst NJT benefited the least from its visual animation, followed by FTW. In both audio-only and audio-visual conditions, preference scores given by male and female participants were not significantly different.

The addition of the visual animation seemed to have helped people to be more confident in stating their preferences. The difference between the most and least preferred water features was larger for the audio-visual condition, in comparison to the audio-only condition.

In the audio-only perception test, four water sounds, namely, CA, DF, FF, and FTW, improved the way participants perceived the environment. The water sound LJT had a neutral effect on people's perception, while NJT was perceived as being detrimental. In the audio-visual condition, all six water sounds appeared to have improved people's perception. In both audio-only and audio-visual conditions, preference scores given by male and female participants were not significantly different.

The average audio-visual perception scores for all water features were higher than their corresponding audio-only scores, suggesting that visual stimuli increased the level of improvement compared to the water sounds alone.

The results showed that the sounds that were given positive scores in the preference test tended to improve people's perception. For the audio-visual conditions, the water features that were positively scored in the preference test (i.e., CA, DF, FF, and FTW), were also highly rated in the perception test. The least preferred water features (i.e., LJT and NJT), improved people's perception the least, yet still positively contributed to changing the environment. These results suggest that, generally, there is a reasonable agreement between perception and preference scores, despite the variation in rankings of the waterscapes, indicating that the water features that tend to be preferred are likely to be beneficial in view of improving people's perception of their environment.

The categorical analysis of the perception scores revealed that CA and FTW significantly improved the sound environment, and no water feature significantly deteriorated the sound environment regardless of their preference levels. In the audio-visual condition, CA, DF, FF, and FTW resulted in a significant improvement in the environment, and no water feature negatively affected people's perception.

The improvement in perception obtained in the audio-only condition was compared to that recorded in the audio-visual condition, using odds ratio. All odds ratios were larger than one, indicating that visual stimuli increased the likelihood of making a positive change in people's perception. The odds ratios showed that there was 1.1 to 2.5 times more chance to receive positive changes in perception when the audio materials were accompanied by visual animations.

Cluster analysis revealed that some people had a clear preference towards the water sounds that they would like to be used to mask irrelevant speech (Cluster 1) while some others perceive most water sounds to be beneficial in masking irrelevant speech (Cluster 2). Roughness was found to have a significant and negative correlation with preference ratings of the water sounds, and no measure was found to have any significant correlation with perception ratings. Furthermore, roughness and pitch strength maintained the same direction of correlation with preference ratings, between the clusters, while the remaining measures tended to have opposing directions of correlation between the two clusters. The sound interference level (SIL) was the only measure to maintain the same direction of correlation with the perception ratings of the water sounds, in both clusters.

In the next chapter, the findings of this chapter are used to examine the effect of masking irrelevant speech with one of the preferred water sounds, on cognitive performance and subjective satisfaction.

CHAPTER 6: COGNITIVE PERFORMANCE AND SUBJECTIVE SATISFACTION

6.1 Introduction

The preference scores reported in Chapter 5 were based on people listening to combinations of speech and water sounds while imagining they were working in an open-plan office. To extend the findings and relate them to cognitive performance, further experiments were necessary. This chapter examines the effectiveness of the preferred water sound in masking irrelevant speech and improving cognitive performance, and people's perception of the work environment, generally, and the sound environment, specifically. The chapter starts by describing the nature of the experiment and the type of water sound, as well as the speech recording, used in it. Cognitive tasks used in previous research are discussed and justifications are given for the tasks selected for this study. Measurement processes are described, including measuring people's performance in selected cognitive tasks in terms of response accuracy and reaction time, under two background noise conditions; a speech-only condition and a masked speech condition. The statistical analyses adopted for analysing the results are explained in detail, followed by findings achieved with regards to improvement in cognitive performance and subjective satisfaction and workload. The chapter ends with a critical discussion of the findings, impact and limitations of the experiment, as well as conclusions of main findings.

6.2 Selected water sound

The water sound selected for this experiment was the 37-jet fountain (FTW) used in the previous experiments. This is in line with the audio-only preference test results presented in Chapter 5 which revealed that FTW was the preferred water sound among the 6 water sounds used in the experiment. Although the inclusion of the visual materials in the previous chapter resulted in higher levels of subjective satisfaction, an audio-only approach was taken in this experiment to allow examining the effect of the water sound per se on people's cognitive performance. The water sound was played from a loudspeaker and its sound pressure level (SPL) was precisely controlled via digital audio processing. This approach allowed for a tighter control over the SPL of the water sound,

in comparison to using a real water feature. Given the higher audio-visual perception score, it was hypothesised that if improvement in cognitive performance could be achieved in this audio-only condition, then a higher level of improvement might be expected in an audio-visual condition. The octave-band frequency spectrum over the frequency of interest (125 to 8k Hz) of the water sound is given in Figure 6.1.

6.3 Speech recording

A high-quality recording of speech was used as a source of irrelevant speech. The speech recording was the same as the one used by Veitch *et al.* (2002) in their study on masking speech in open-plan offices. The original recording consisted of 17 minutes of almost continuous speech of a single female voice speaking at a realistic speech level. The speech comprised one-sided dialogues simulating one side of telephone conversations, represented by the voice of an actress reading scripts of telephone conversations, in which she called job candidates to arrange for interviews, made internal arrangements for new employees and made personal social calls.

The equivalent sound pressure level of the speech recording $L_{Aeq,17 \text{ min}}$ was calibrated to 48 dB, in line with recommendations from previous research discussed in Chapter 3. The spectrum of the speech recording was measured in octave bands from 125 Hz to 8 kHz at 1-second intervals over the length of the speech recording, i.e., 17 minutes. For each frequency band, the sound pressure level at each 1-second interval was averaged logarithmically to generate the energy averaged spectrum of speech. The result is shown in Figure 6.1 alongside the reference unisex speech spectrum recommended by BS EN ISO 3382-3 (2012). The largest variation of SPL happened at mid frequencies, with 500 Hz having the largest standard deviation of 22.2 dB. The higher frequencies showed a relatively smaller variation in SPL with the 8-kHz band having the smallest standard deviation of 13.9 dB. The speech recording used in this study appeared to have a close match with the reference speech spectrum at most octave bands except for the 8-kHz octave band. The SPL of the speech recording is approximately 10 dB higher than the reference speech level at 8-kHz octave band. This is largely due to the fact that the reference speech spectrum is a unisex speech spectrum, while the recording used in this study is from a female speaker. Secondly, in the speech recording, there were a few occasions when a telephone rang that could have increased the SPL of the higher frequencies.

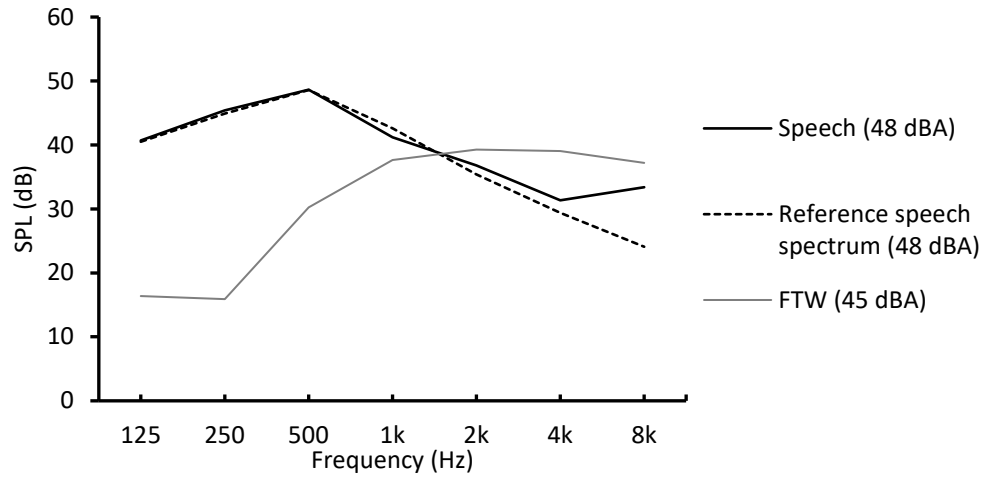


Figure 6.1 The spectrum of the 37-jet fountain (FTW) and the energy averaged spectrum of speech (averaged over 17 minutes at 1-second intervals). The dashed line is the reference unisex speech spectrum recommended by BS EN ISO 3382-3 (2012).

Figure 6.2 is the $L_{Aeq,5s}$ of the speech recording, integrated at 5-second intervals over the length of the recording, showing the variation of SPLs over time. The relatively wide white gaps are the pauses between the different telephone conversations (6 different conversations in total).

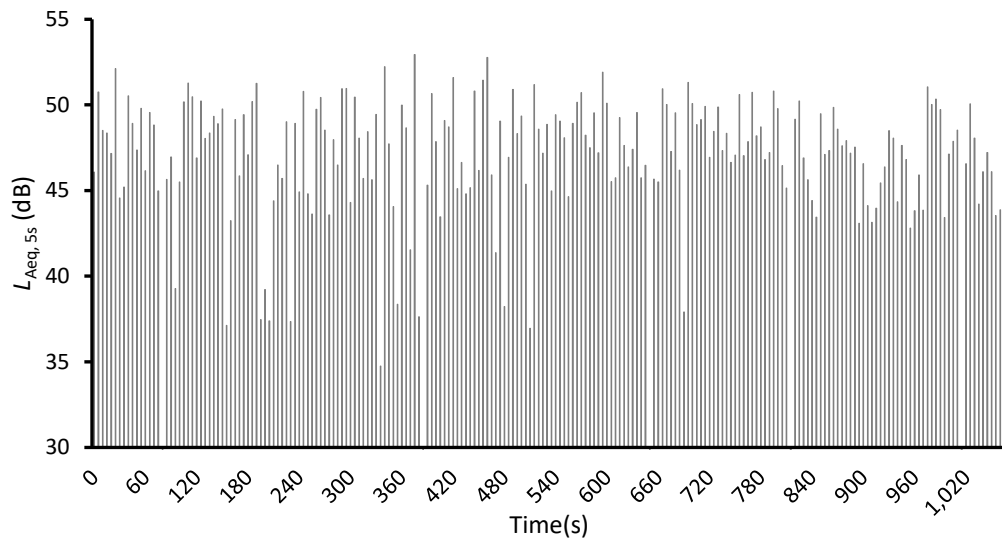


Figure 6.2 A-weighted equivalent sound pressure level $L_{Aeq,5s}$ of the speech recording used in the current study, for the duration of the entire speech recording (an $L_{Aeq,17m}$ of 48 dB).

6.4 Participants

Forty-eight participants (24 males, 24 females) who reported a normal hearing ability took part in the experiment. These were staff members and post-graduate students from Heriot-Watt University. The age distribution of participants was between 24 and 64 yr with an average of $M = 35.48$ yr ($SD = 10.86$ yr). Out of the selected sample, 23 participants were native English speakers (10 males, 13 females), and the remaining 25 participants (14 males, 11 females) were non-native speakers who spoke English fluently¹. Figure 6.3 shows the differences in age distribution of participants between males and females, as well as between native and non-native speakers.

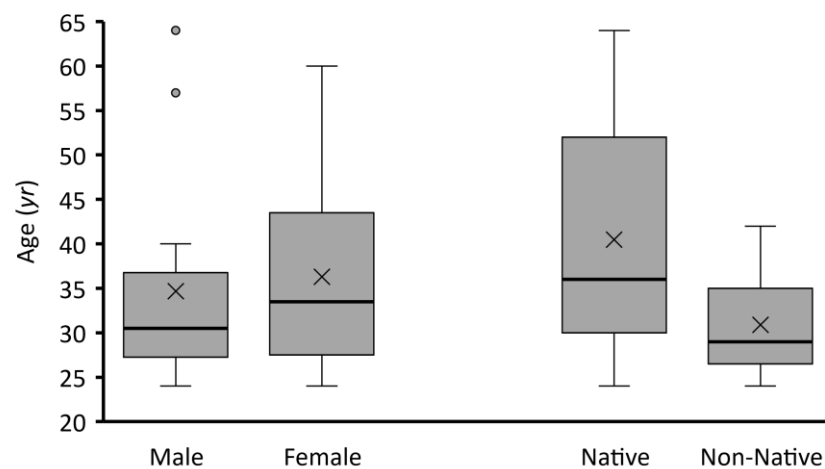


Figure 6.3 Age distribution of participants in the cognitive task performance experiment. The tinted boxes represent the interquartile range (i.e., 50% of scores) with the top and bottom of the tinted box representing the upper and lower quartiles, respectively. The bold horizontal lines inside the boxes are medians, and the cross marks (×) are the means. Small circles represent unusual cases and the whiskers show the top and lowest 25% of scores.

The boxplots show that the average age of males and females were similar but the distribution was more widely dispersed for females. There were two male participants whose ages were unusually higher, which might have slightly biased the mean average age of male participants. The difference in age is more apparent between native and non-native speakers, with native English speakers being older and the age distribution being wider in comparison to non-native speakers. Understanding these differences in age distribution is important as it will help make a better interpretation of results, should there

¹ Under the current UK laws, it is a requirement for non-native people to meet a certain level of proficiency in English to be able to study or work in the country. The current minimum requirement is IELTS 6.5.

be statistical differences between males and females and/or between native and non-native speakers. Statistical differences between these groups could in fact be partially attributed to differences in their age distributions. At the end of the experiment, participants received a 5-pound Amazon voucher, as a token of gratitude for taking part in the experiment.

6.5 The open-plan space

There was no space in the acoustic laboratory to allow for this test to be carried out inside it. Therefore, it was decided to perform the experiment in a quiet room where the background noise was relatively low with no impulsive or distracting noises. The tests were carried out in Room G.33 of the Edwin Chadwick Building at the Edinburgh Campus of Heriot-Watt University. The space was a small room with an area of 13.5m^2 ($3.75\text{L} \times 3.60\text{W} \times 2.5\text{H m}$) as shown in Figure 6.4 and Figure 6.5. A photograph of the space is shown in Figure 6.6. Although this room does not represent an open-plan office, it was proven difficult to find a larger space, and lack of equipment allowed only two participants to be tested at a time. This is a limitation of the study that could affect its finding, as a larger space with more participants would have created a more realistic open-plan space. Having said that, the objective of the experiment was to investigate the effect of masking irrelevant speech on task performance, and therefore, the effect of the space itself on the cognitive performance might not be as important as the effect of the background noise.

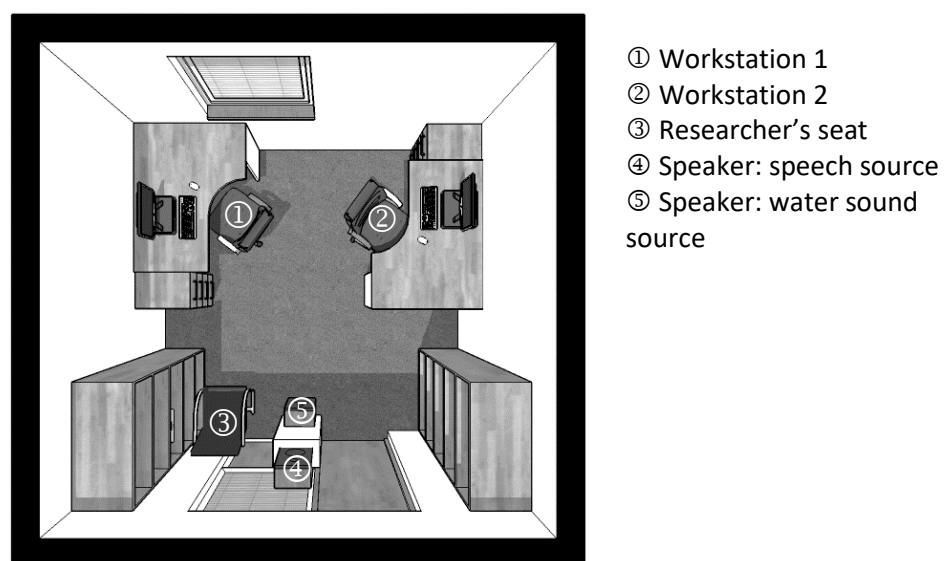


Figure 6.4 A sectional perspective of the test room showing the arrangement of furniture, workstations and speakers.

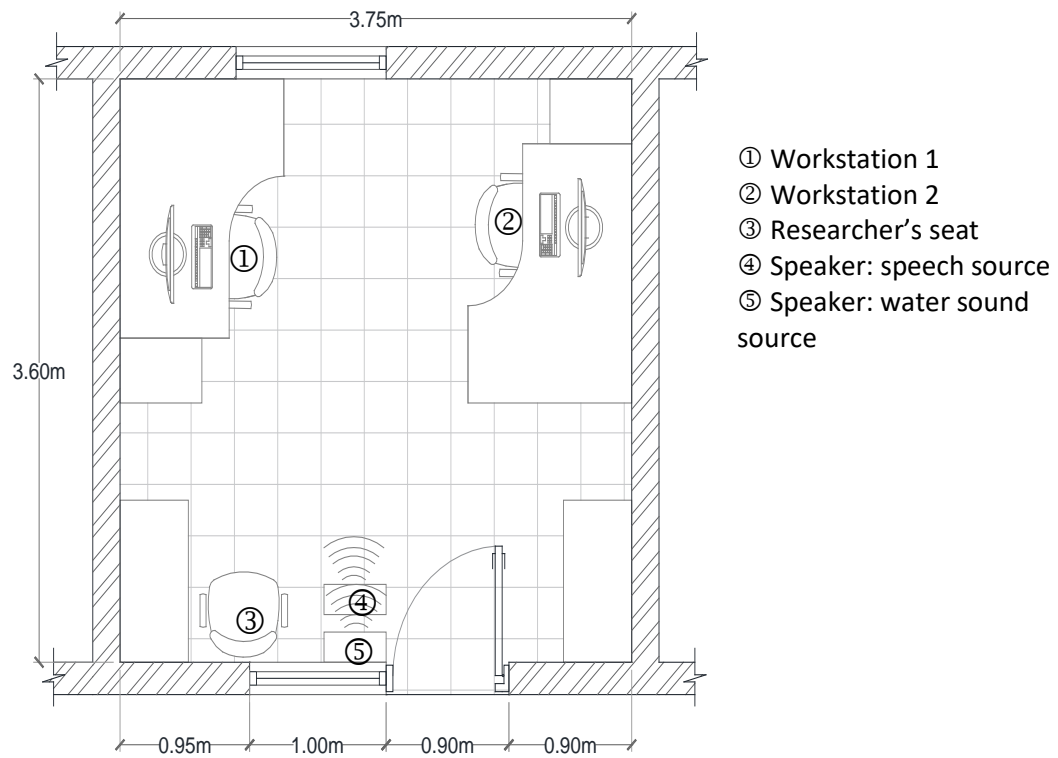


Figure 6.5 Test room in which the task performance experiment was carried out. Dimensions are in metres.



Figure 6.6 A photograph of the test room showing the workstations.

The speech was played from a speaker (item 4 in Figure 6.4 and Figure 6.5) placed on a stand making the speaker's height 1.2 m, which was measured at the midway point between the woofer and the tweeter of the speaker and the ground. This source's height

represents speech from a person seated. The water sound was played through another speaker (item 5 in Figure 6.4 and Figure 6.5) which was placed on the ground in front of the speech source.

The room had a window (dimensions, $1.10W \times 1.15H$ m) with a blind which was turned halfway down to allow daylight to enter the space while preventing direct sunlight which would have caused glare on the computer screens. The artificial lighting was kept on ensuring a uniform distribution of light across the workstations. Having both natural light and artificial light was based on results of a post occupancy survey (Roche *et al.*, 2000), which showed that the majority of people working in open-plan offices to prefer to work in natural light or a combination of natural and artificial lighting.

The reverberation time of the test room was measured at each third-octave-band from 100 Hz to 8 kHz. The frequency averaged reverberation time was 0.40 s. Two source positions, each with three microphone positions, were used in the measurements of the reverberation time, and the results are graphically shown in Figure 6.7. Reverberation time measurements were carried out in accordance with BS EN ISO 3382-2 (2008). The profile of the reverberation time is not typical of an open-plan office, mainly due to the small volume of the space where the experiment was carried out. However, reverberation time is not a critical factor in dictating the acoustic performance of open-plan offices (Virjonen *et al.*, 2009). Hence the deviation from a typical reverberation time profile is unlikely to have affected the results of this experiment.

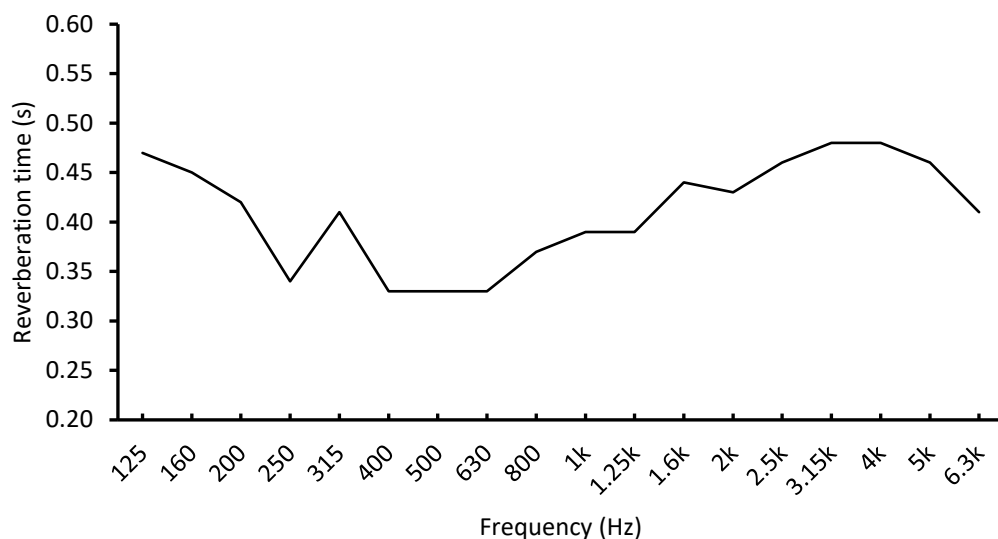


Figure 6.7 One-third octave band reverberation time (in seconds) of the test room.

6.6 Background noise conditions

The tasks were carried out under two background noise conditions: a speech-only condition and a masked speech condition. During the speech-only condition, the speech recording described in Section 6.3 was played from a speaker at 1.2 m height (midway point between the woofer and the tweeter), as shown in Figure 6.4 and Figure 6.5. The speech signal was calibrated to have an $L_{Aeq, 17 \text{ min}}$ of 48 dB at both workstations 1.2 m above the ground (ears' height of a person seated). For the masked speech condition, the speech signal was kept as in the speech-only condition, but a water sound (FTW) was played from a speaker placed on the ground underneath the speech source, at 0.3 m height (midway point between the woofer and the tweeter), as shown in Figure 6.4 and Figure 6.5. The water sound was calibrated to have an $L_{Aeq, 20 \text{ s}}$ of 45 dB (preferred level found in Chapter 4) at both workstations, 1.2 m above the ground.

6.7 Cognitive tasks

People perform various tasks in open-plan offices from writing, reading, data entry, short and long-term memory tasks, logical reasoning and mathematical tasks. These tasks require different cognitive abilities. Investigating all these cognitive abilities would be costly and result in overly extended tests in which participants would feel exhausted. This, in turn, would negatively affect the findings of the study. On this ground, it was decided to investigate only a limited number of cognitive tasks. An extensive literature review was carried out to identify potential cognitive tasks, and in doing so, two selection criteria were used. The first criterion was the length of the task, and the second one was whether the task had been affected by background noise conditions in previous research. The overall length of the experiment was set to be no longer than 45 minutes, to avoid participants being too fatigued. The tasks that have been used in previous research mainly focus on using different types of memory. Serial recall task is a short-term memory-based task which requires the participant to memorise information presented on a screen for only a few seconds, while reading tasks require memorising information for a longer period. Writing tasks require the activation of long-term memory.

A list of cognitive tasks used in previous studies is given in Table 6.1. Only conditions which included a masking noise in the background noise are reported in the table. The silent condition was disregarded as it is well documented that people perform better in the absence of any disturbing background noise including irrelevant speech.

Table 6.1 Cognitive tasks used in previous research in the context of performance and masking background noise. (Continued)

#	Study	Task	Description of Task	$p < .05$
1	Liebl <i>et al.</i> (2016)	<i>Serial Recall</i>	Numbers from 1 to 9 presented on a screen in a specific order. Participants would then need to memorise the numbers and their order in the presentation.	Yes
2	Keus van de Poll <i>et al.</i> (2014)	<i>Writing task</i>	Participants were asked to write a short essay about a subject. Measurements included the number of words written in each essay, the number of deleted words and the number of times participants took a pause for more than 5 seconds.	No
3	Haapakangas <i>et al.</i> (2014)	<i>Serial recall</i>	Same as Liebl <i>et al.</i> (2016).	No
		<i>Operation span task</i>	Equations and words presented in turns. First, an equation appeared on a computer screen and the participant would decide whether the expression was true or false. After that, a word appeared on the screen and participant would need to memorise the word, to be recalled later.	Almost ($p = .07$)
		<i>N-back task</i>	Sequences of letters presented on a computer screen, and the participant would decide if the letter is identical to a target letter (0-Back), the one immediately preceding it (one trial back, 1-Back) or the letter presented two trials back (2-Back).	Accuracy: no Reaction time: yes
4	Jahncke <i>et al.</i> (2013)	<i>Text memory task</i>	A three-page long text to be read by participants in 6 minutes. After 30 min subjects were asked to recall as much detail as they could about the text.	No
		<i>Word memory task</i>	Participants were shown words from lists with a number of words in each list. Participants would then recall words from the most recent list and ignore words from earlier lists.	No, between STI .34 and .71
		<i>Information search task</i>	Table containing 20 rows and 7 columns. For each row, an object was presented and each column described one aspect of the object. Participants were asked to find the object that met a set of criteria, by using 3 of the 7 columns.	No, between STI .34 and .71
4		<i>Math task</i>	Triple-digit numbers, to be added by participants.	No between STI .34 and .71

Table 6.1 Cognitive tasks used in previous research in the context of performance and masking background noise. (Continued)

#	Study	Task	Description of Task	$p < .05$
		<i>Word fluency tasks</i>	Consisted of two tasks: semantic and phonemic fluency tasks. For the semantic fluency task, participants were asked to write as many examples from a given semantic category as possible within 60 s. In the phonemic fluency task, participants were presented with a letter and instructed to generate as many words beginning with that letter as possible within 60 s.	No
5	Jahncke <i>et al</i> (2011b)	<i>Proactive interference</i>	Similar to “Word memory task” in Jahncke <i>et al.</i> (2013).	Yes
		<i>Response inhibition</i>	Digits from 1 to 9 were presented repetitively and participants were instructed to respond to all digits except number 3.	No
		<i>Flanker task</i>	Participants were presented with strings of five letters composed of H and S (e.g., HHHHH, HHSHH). They had to respond to the central letter by pressing “h” or “s” on the keyboard.	No
		<i>Logical problems</i>	Participants were presented with logical problems and were told to answer whether the conclusion drawn was true or false (e.g., All A are B. Some B are C. Conclusion: some A is C. Is this conclusion true?).	No
		<i>Operation span</i>	Same as Operation span task in Haapakangas <i>et al.</i> (2014).	No
		<i>Reading comprehension</i>	Ten short texts presented. Participants were required to answer one question per text out of four alternative answers.	No
		<i>Serial recall</i>	Similar to other serial recall tasks but letters and numbers were included.	No
		<i>Shifting</i>	Number-letter pairs (e.g., 3E) presented in a clockwise rotation around all four quadrants of a computer screen. When the pair was presented in the top two quadrants, participants were to indicate whether the number was even or odd, and when presented in the bottom two quadrants, they were to indicate whether the letter was a vowel or a consonant.	No
6	Haapakangas <i>et al.</i> (2011)	<i>Updating</i>	Lists which contained 2-digit numbers presented and participants were required to recall the three smallest numbers of each list.	No
		<i>Serial recall</i>	Similar to other serial recall tasks.	Yes
		<i>Creative thinking task</i>	The subjects were instructed to write down as many alternative uses for an object as possible in 5 minutes.	No
		<i>Proofreading</i>	Texts containing typographical and grammatical errors to be proofread by participants.	No

Table 6.1 Cognitive tasks used in previous research in the context of performance and masking background noise. (Continued)

#	Study	Task	Description of Task	$p < .05$
7	Schlittmeier and Hellbrück (2009)	<i>Serial recall</i>	Similar to other serial recall tasks.	Yes
8	Haka <i>et al.</i> (2009)	<i>Serial recall</i>	Similar to other serial recall tasks.	Yes
		<i>Operation span task</i>	Similar to “operation span task” in Haapakangas <i>et al.</i> (2014).	Yes
		<i>Reading comprehension task</i>	Participants were presented with a three-sentence paragraph, which described relationships between real and invented words. Participants were then presented with statements and they had to decide whether the statements were true or false.	Almost ($p = .05$)
		<i>Proofreading</i>	Texts containing typographical and grammatical mistakes to be proofread by participants.	No
9	Venetjoki <i>et al.</i> (2006)	<i>Reaction time</i>	A number appeared on a screen with a random delay ranging from 1 to 4 s. Participants were instructed to react to the stimulus by pressing the number on the keyboard as soon as they detected the stimulus.	No
		<i>Subtraction task</i>	Similar to the above task but participants had to subtract numbers presented on the screen from number 9.	No
		<i>Proposition task</i>	Whole sentences (e.g., 7 is larger than 8) presented on a screen, one word at a time. Participants had to decide if the sentences were logical or not.	No
		<i>Vigilance task</i>	Letters were presented on a screen and participants were instructed to press a button on the keyboard for only three target letters.	No
		<i>Stroop effect</i>	The words ‘red’ written in blue and ‘blue’ written in red were presented to participants and they were instructed to react to the colour of the printing by pressing corresponding keys.	No
		<i>Reading comprehension</i>	Three to four-page written text was presented to participants and they were asked to answer a set of questions concerning the text.	No
		<i>Proofreading</i>	A four-page text was given to participants and they were asked to identify as many typographical and grammatical mistakes as possible.	Yes

It is clear that there is a discrepancy in results reported by different studies. For instance, serial recall tasks have been reported to be negatively affected by the background noise condition in some studies (Haka *et al.*, 2009; Schlittmeier and Hellbrück, 2009; Haapakangas *et al.*, 2011; Liebl *et al.*, 2016), while other studies (Jahncke *et al.*, 2011b;

Haapakangas *et al.*, 2014) found no difference in performance of serial recall tasks between masked and unmasked speech conditions. Ebissou *et al.* (2015) argue that STI is not the only factor affecting performance in serial recall tasks. The study states that, in addition to the STI, cognitive performance can strongly depend on participants, especially if a relatively small sample size is used, which is often the case in many relevant studies. This could partly explain the discrepancy between results reported by different studies on speech intelligibility and cognitive task performance in serial recall tasks.

The contradiction extends to operation span tasks which have been reported by Haka *et al.* (2009) to be affected by noise conditions, while Jahncke *et al.* (2011b) and Haapakangas *et al.* (2014) found no statistically significant differences in performance of such tasks under masked and unmasked speech conditions. The results obtained for reading tasks, on the other hand, seem to be quite consistent across the studies. Tasks which involved reading showed not to be affected by the background noise conditions in four studies (Venetjoki *et al.*, 2006; Haka *et al.*, 2009; Jahncke *et al.*, 2011b; Haapakangas *et al.*, 2014). Drawing conclusions on the basis of currently available literature is difficult, but it would appear that tasks that require using the short-term memory, such as serial recall, are more susceptible to be negatively affected by the change in the background noise condition (i.e., cognitive performance tends to be higher in a masked speech condition, compared to a speech-only condition).

The nature of the tasks reported in Table 6.1 can be classified into writing, reading, short-term and long-term memory tasks, proofreading, information search and updating, mathematical tasks and logical reasoning. Attempts were made to cover most of these tasks without making the experiment too long. A proofreading task was excluded as non-native English speakers were planned to participate and they were expected to perform poorer than the native speakers in this task, which would have unnecessarily complicated the analysis of the results. Tasks that involved writing and creative thinking were also excluded because of the lack of a widely accepted method of carrying out and assessing such tasks. People's performance in writing tasks could be assessed from the proofreading's point of view, but writing is a more complex process that involves creative thinking which cannot be assessed easily in an objective way, and this was therefore disregarded. After careful consideration, four tasks, which covered different types of cognitive abilities, were selected, namely, serial recall, one-back, information matching and reading comprehension. These are described in detail in the following sub-sections.

6.7.1 Serial recall task

The serial recall is a typical task investigating the effect of irrelevant speech on cognitive performance. The task relies on short-term memory and requires participants to memorise a nine-digit figure for a period of no more than 30 seconds. This test was selected to be included in the experiment, as it has been used a number of times in previous studies and is likely to be affected by the background noise.

In the task, digits from 1 to 9 were presented in the centre of a computer screen in a specific order, one after another. Participants had to recall the digits and their orders, after all nine digits had been presented. Each digit was presented once in a sequence and remained on the screen for 1 second. Between each two digits, there was a 1s gap (blank screen). Participants recalled the digits by clicking on a 3×3 array on the screen as shown in Figure 6.8. Participants were instructed to avoid guessing and therefore an (\times) button was provided on the screen in case a participant could not remember a digit in a certain serial position. The task contained 5 sequences and the number of digits correctly recalled in their positions was recorded and averaged out over the 5 sequences. There was no time restriction for participants to recall the sequences, but the time taken by participants to recall each sequence of digits was measured and averaged out over the 5 sequences. Two sets of this task were prepared, each to be used in the speech-only and the masked speech condition. The task lasted around 3 minutes. Instructions and data used in the five trials of the serial recall task are provided in Appendix F.

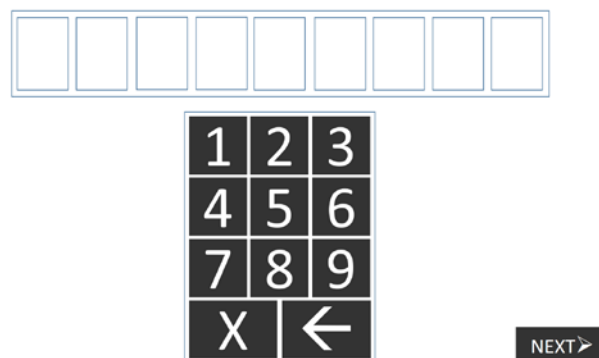


Figure 6.8 A screenshot of the interactive array of digits used in the serial recall test.

6.7.2 One-back task

The n -back task is an example of a working memory task which involves using short-term memory and manipulation of remembered information (Owen *et al.*, 2005). The task requires participants to observe a sequence of verbal or nonverbal stimuli and indicate if the currently presented stimulus is identical to the one presented “ n ” trials back, where n

is an integer (e.g., 1, 2, and 3). The task has been widely used in many human studies investigating the working memory process (Owen *et al.*, 2005). The task used in this study is similar to the n-back task used by Haapakangas *et al.* (2014), which included, 0-back, 1-back and 2-back tasks. However, in the current study, only 1-back was included, as 0-back was seen to be too easy and 2-back was relatively difficult and would have made the test last longer.

The task consisted of a set of letters randomly presented in the middle of the computer screen. Each letter remained on the screen for 1 second, followed by a blank screen with a green (✓) sign on the right side and a red (×) sign on the left side of the screen. Participants were instructed to click on the green (✓) if the currently presented letter was identical to the one immediately preceding it (i.e., one trial back), or to click on (×), otherwise.

The test contained 31 letters (30+n); one-third of them were target letters requiring a positive response from the participant (i.e., click on ✓). Both response accuracy (%) and reaction time (s) were measured. Two sets of the task were prepared, each to be used in one of the two background noise conditions. The task lasted approximately 1 minute. Instructions and data used in the one-back task are provided in Appendix G.

6.7.3 Information matching task

This task is similar to a typical data entry task, and therefore, it was decided to include it as a part of this experiment. This task requires participants to search through a table of information while memorising and updating information in accordance with set criteria. In principle, this task is similar to the “information search task” used in a previous study (Jahncke *et al.*, 2013), but the approach taken here was different.

A table of information was presented to participants. The table contained nine rows ordered from 1 to 9, each with three columns as shown in Figure 6.9. Each row described a button through using three pieces of information. The first piece (column 1) described which letter the button should contain (A to I). The second piece (column 2) described the digit (1 or 2) that the button should have and the last column provided information regarding the colour of the button (yellow or blue). All possible combinations of buttons were given in a 9×4 array of buttons and participants were instructed to fill in a 3×3 array matrix on the top of the screen by clicking on the appropriate buttons that met the description given in the table. For instance, the first row of the table in Figure 6.9 describes a yellow button that has the letter A and number 1 in it. Participants would need

to find that button in the array of buttons and click on it to be transferred to the first cell of the 3×3 array. A completed example of the task is given on the right-hand side of Figure 6.9.

Participants repeated the task five times, each with a different table of information. The response accuracy (%) and the time (in seconds) taken by them to complete the task were measured and averaged out for the 5 trials. Two sets of the task were prepared, each to be used in one of the background noise conditions. The task lasted for approximately 4 minutes. Instructions and data used in the five trials of the information matching task are provided in Appendix H.

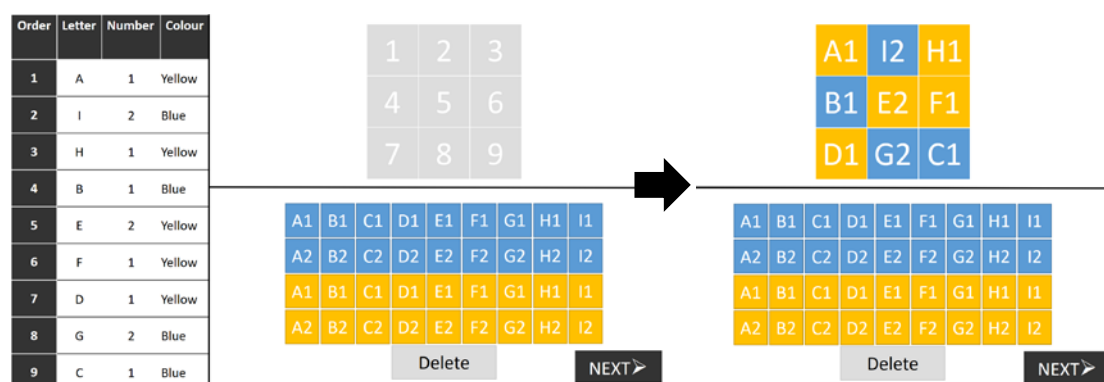


Figure 6.9 A screenshot of the information matching task.

6.7.4 Reading comprehension task

Reading comprehension is another typical office task. The conventional approach of measuring people's performance in this task involves presenting a relatively long (1 to 3 pages) text to participants followed by multiple choice or true/false questions. Hannon and Daneman (2001) developed a new tool to measure component processes of reading comprehension, that has been reported to be easier to administer and to have high predictive power in comparison to the conventional methods. This new tool comprises a three-sentence paragraph that describes the relationship among a set of real and imaginary terms. For instance, a three-sentence paragraph used in this study was as follows: "A MIRT resembles a Lion but is larger and eats more. A COFT resembles a Chicken but is smaller and eats more. A FILP resembles a COFT but is smaller and eats more." In the paragraph, terms MIRT, COFT and FILP are nonsense words invented for the study, while Lion and Chicken are meaningful words. The paragraph describes two features of those five terms, which are size and diet. Participants were expected to find linear orderings using the information given in the paragraph. For example, for the feature "size"

the linear ordering is MIRT > Lion > Chicken > COFT > FILP. However, the fact that a Lion is larger than a Chicken was not specifically given in the text, participants would need to access their prior knowledge to make this ordering. For the feature “diet” participants were expected to draw two linear orderings, MIRT > Lion > Chicken, and, FILP > COFT > Chicken, through using their prior knowledge that a Lion eats more than a chicken. Participants were then presented with true/false statements that assessed their reading comprehension abilities on four component processes. *Text memory statements* (e.g., A MIRT is larger than a Lion) tested memory for information that was directly given in the text and did not require activating the prior knowledge. *Text inferencing statements* (e.g., A FILP is smaller than a Chicken) tested implicit information that could be concluded through combining information given explicitly in the text (e.g., A COFT resembles a Chicken but is smaller and eats more. A FILP resembles a COFT but is smaller and eats more). *Knowledge access statements* (e.g., A Lion is larger than a chicken) tested prior knowledge information without the need to study the text, i.e., no input from the text was required to assess these statements. *Knowledge integration statements* (A MIRT eats more than a Chicken) tested implicit information that could be assessed through accessing the prior knowledge (A Lion eats more than a chicken) and integrating this knowledge with the information given explicitly in the text (A MIRT eats more than a Lion). The components of the task allow for predicting reading comprehension performance on a global scale (using all four component processes) and on each of the comprehension components.

The reading comprehension task was originally developed in Canada and contained terminologies that were common or native to North America (e.g., robin, and blue jay) but less common in the UK, and potentially misleading. Therefore, a modified version of the test was developed, in which more simplified words (e.g., lion, chicken) were used.

The modified reading comprehension task included 20 true/false statements (identical to the original study (Hannon and Daneman, 2001) in terms of structure) . Six statements were Text memory, two statements were Text inferencing, eight statements were Knowledge integration, and the remaining four statements were Knowledge access statements. There was only one level of difficulty for Text memory and Text inferencing statements, while Knowledge integration had three levels of difficulty (low, medium, and high) and Knowledge access had two levels of difficulty (low, and high). The three-sentence paragraph was presented in the middle of the computer screen one sentence at a time. Participants controlled how long a sentence would remain on the screen as there

was no time restriction on reading and studying the sentences. Once they studied a sentence, they would then click on a “Next” button on the screen to see the next sentence. After they had seen all three sentences, the statements appeared on the screen one at a time in a random order. For each statement, participants had to decide whether that statement was true (✓) or false (×). Half of the statements (i.e., 10 statements) were true and the remaining half were false statements requiring participants to give a negative response. The response accuracy (%) was measured for the overall reading comprehension task, the individual components and their corresponding difficulty levels. Furthermore, the times taken by participants to read the paragraph and to respond were measured. Two sets of the test were prepared, each to be used in one of the background noise conditions. The full list of statements and paragraphs used in this task can be found in Appendix I.

6.8 Questionnaires

The questionnaire used in this part of the study consisted of four sections. The first section was on background information of participants such as age, gender, nationality, and sleep during the preceding night. The second section was the noise sensitivity questionnaire, assessed using the shorter version of the noise sensitivity questionnaire (NoiseQ) developed by Schütte, Sandrock, *et al.* (2007), which consists of 12 items divided into 3 subscales, namely, sleep, habitation and work, with 4 items in each subscale. Only the subscale “work” was used in the current study, and hence, 4 items were included in this section of the questionnaire (Table 6.2). Participants stated their level of agreement with each item on a 4-point numerical scale with 1 representing “strongly disagree”, and 4 representing “strongly agree”. The answer to each question was then quantified from 0 to 3 and used to calculate the average noise sensitivity score. A score of less than 1.11 is considered as not being sensitive to noise, while a score greater than 1.63 is considered as being sensitive to noise (Schütte, Marks, *et al.*, 2007).

The third section was the subjective satisfaction questionnaire and the last section concerned the subjective difficulty of the tasks, which was designed to measure how difficult people found each task in the two background noise conditions. These are described in the next subsection.

The background information and noise sensitivity questionnaires were completed only once at the beginning of the test. The satisfaction questionnaire was completed twice, at the end of the last experiment in each of the two background noise conditions. The

subjective task difficulty questionnaire was completed at the end of every task in each background noise condition.

6.8.1 Subjective satisfaction questionnaire

Peoples' perception and their subjective satisfaction of the work environment are important. An improvement in cognitive performance might not yield the required result if it is not accompanied by an increased level of subjective satisfaction. Based on empirical findings from Chapter 5, it was hypothesised that the water sound would increase people's satisfaction with the work environment. This hypothesis was investigated by using a short satisfaction questionnaire. The descriptions and the scale used for each item included in the satisfaction questionnaire are given in Table 6.2.

Table 6.2 List of items/questions used in the questionnaire in the task performance experiment.

Sections	Items/questions	Scale
1 Background information	What is your gender?	Male/female
	What is your age?	N/A
	What is your nationality?	N/A
	How many hours did you sleep last night?	N/A
2 Noise sensitivity	I need a quiet environment to be able to carry out new tasks.	
	When people around me are noisy, I find it hard to do my work.	1-4: strongly disagree -
	I perform significantly worse in noisy environments.	strongly agree
	I need peace and quiet in order to carry out a difficult task.	
3 Subjective satisfaction	<i>Item1</i> : The sound environment was pleasant.	
	<i>Item2</i> : My attention was drawn to the sound environment.	
	<i>Item3</i> : The sound environment helped me to concentrate on the tasks.	0-10: strongly disagree -
	<i>Item4</i> : The speech disturbed me.	strongly agree
	<i>Item5</i> : I could have meetings in my office without distracting others.	
	<i>Item6</i> : I could work uninterrupted for long periods.	
	<i>Item7</i> : The noise in my office would not be distracting.	
	<i>Item8</i> : I could easily have confidential conversations.	
4 Task difficulty	What percentage of the speech did you understand?	0% to 100%
	On a scale of 0 to 10, how difficult the task was?	0-10: extremely easy -extremely difficult

The satisfaction questionnaire included items each designed to capture one aspect of people's perception of the sound environment. The pleasantness and distraction level of the sound environment, the ability to concentrate and the disturbance caused by irrelevant

speech were aspects that were measured by the questionnaire. Items 1, 2, and 3 were adapted from Haapakangas *et al.* (2014), due to their relevance to the current study. Items 5, 6, 7, and 8 were future-oriented as they required the participant to imagine how would they respond to the items if they had worked in that environment for a long period of time. The four items were taken from Veitch *et al.* (2002) and concerned the possibility of having a meeting without distracting others, the distraction level, and the possibility of having confidential conversations. An 11-point numerical scale was used in this section of the questionnaire, where 0 stood for “strongly disagree”, and 10 stood for “strongly agree”. The main aim of the questionnaire was to examine the likely effects that masking background speech with a water sound had on people’s satisfaction and perception of their work and sound environment, and small differences are more easily detectable in an 11-point scale than, for example, a 5-point Likert scale. People were also asked to subjectively rate the speech intelligibility level as a percentage, where 0% stood for not intelligible at all, and 100% stood for perfectly intelligible.

6.9 Procedure

Two participants at a time took part and the test, which lasted for around 45 minutes, during which the researcher remained in the room but did not intervene with the test. Participants were invited into the room and seated at the dedicated workstations on upholstery chairs back facing each other and they were asked not to communicate with each other. They were asked to confirm that they did not have any known hearing difficulties by ticking a box on the computer screen. The participants started the test by filling out a short questionnaire on background information and their sensitivity to noise (refer to sections 1 and 2 of the questionnaire in Table 6.2). Participants were briefed about the nature of the experiments and the number of tasks and conditions they were expected to take part in. They were also informed that completing the test was not compulsory and therefore, they could leave the room at any time during the test, without providing any explanation. Participants then started a practice session which included all the four tasks in the order used for the actual test. Detailed instructions were given on the screen and participants were given verbal instructions when they required further clarification. The instructions given to participants can be found in Appendix E. The practice session was carried out in a silent condition (i.e., no speech nor water sound was played). Participants could repeat the practice session for as many times as they wanted, but they were not allowed to start the real test until both participants had completed the practice session. To account for the order effect, the order in which participants carried

out the tasks was counterbalanced. Participants were divided into 4 groups, each containing 12 participants. Each group did the tasks in a particular order as shown in Table 6.3. The order of the tasks between the speech-only and masked speech conditions was not changed.

Table 6.3 Order of tasks used in the task performance experiment.

Groups	1 st task	2 nd task	3 rd task	4 th task	No. of Participants
Group 1	SR	IM	OB	RC	12
Group 2	RC	OB	IM	SR	12
Group 3	OB	RC	SR	IM	12
Group 4	IM	SR	RC	OB	12

SR = Serial recall task. IM = Information Matching task. OB = One-back task. RC = Reading comprehension task.

Once a task was completed by a participant, the task difficulty questionnaire appeared on the computer screen, which asked the participant to rate the difficulty of the task on an 11-point numerical scale before the next task would start. Once all four tasks were completed by a participant, the satisfaction questionnaire would appear on the computer screen. Once both participants in the room completed the tasks and the questionnaire, the researcher would announce the end of that condition and would change the background noise. Participants would then be asked to carry out the same set of tasks (with different contents) and questionnaire, in the new background noise condition. Half of the participants carried out the speech-only condition first, and the remaining half carried out the masked speech condition first. Participants were allowed to have a short break between the two background noise conditions.

Participants were instructed to ignore the background noise at all times and concentrate on carrying out the tasks. They were asked to be as accurate as possible but the reaction time (i.e., speed) was not mentioned. Both participants had to complete the tasks and questionnaire in one condition before moving to the second condition. At the end of the test, participants were shown their scores for the cognitive tasks in both conditions, and they were given a 5-pound Amazon voucher as a token of gratitude.

6.10 Equipment and software

Cognitive tasks were produced using Microsoft®'s Visual Basic for Applications in PowerPoint. A macro-enabled interactive PowerPoint presentation was prepared for each

participant (i.e., 48 presentations), that allowed the participants to perform the four cognitive tasks.

Digital audio processing was performed using Studio One 3 audio production software (PreSonus Audio Electronics) installed on a laptop computer (Dell XPS L502X). A two-channel sound file was prepared that included the speech signal on the right channel, and the water sound on the left channel. Each channel of the audio file was then connected to one of the speakers. The original speech recording was a mono audio file, while the water sound was a binaural recording. Therefore, only the left channel of the water sound was used to create the two-channel sound file. The water sound and speech were calibrated using a Brüel and Kjær handheld sound analyser, Type 2250.

The sound file was played from the laptop computer, which was connected to an external M-Audio USB sound card. The external sound card was connected to an A 28 J.E.Sugden two-channel sound amplifier. Each channel of the amplifier was connected to a loudspeaker, KEF Coda III, Type SP3016 (maximum amplifier power: 50 W, Frequency range: 65 Hz to 20 kHz).

Two Hewlett-Packard (HP EliteDesk 800 G1 SFF) workstation computers were used to run the tasks. All tasks and questionnaires were presented on the computer screen (HP EliteDisplay E231) and required participants to react and respond through clicking on a computer mouse (no keyboard used).

6.11 Statistical analysis

6.11.1 Multivariate approach

Task performance can be defined by both response accuracy (RA) and reaction time (RT). Ignoring one of these could result in misleading conclusions as there is often a trade-off between the two. Because both RA and RT represent performance, a multivariate analysis of variance (MANOVA) was adopted in the statistical analysis of the measured data. This approach allowed for analysing both RA and RT as two dependent variables representing performance. MANOVA takes account of the relationship between variables, allowing for investigating whether groups (e.g., gender) are different along a combination of dimensions (Field, 2013).

Lamb and Kwok (2016) used a combined effect of the reaction time and the response accuracy as a measure of performance. They added up the response accuracy score and the reaction time score (reversed) to generate a single-number score to account for the

time/accuracy trade-off. Using this approach in this study would have allowed for using simpler statistical models (e.g., ANOVA). The drawback of this approach, from a statistical point of view, is that it assumes an equal effect of reaction time and response accuracy, which is not necessarily true. For some tasks, the response accuracy might be much more important than the reaction time. Furthermore, this approach does not allow for analysing the interaction between the reaction time and the response accuracy. As a result, a multivariate approach was favoured in this study over a univariate approach.

A mixed design multivariate analysis of covariance (MANCOVA) approach was adopted as same participants were used in both background noise conditions. The effects of gender and nationality (whether a participant was a native English speaker), were also of interest to this study. Therefore, they were included in the statistical analysis as two between subject variables. Previous research suggests that individuals who are sensitive to noise are more susceptible to be affected by background speech both in terms of cognitive performance and subjective satisfaction (Haapakangas *et al.*, 2014). Therefore, the noise sensitivity scores of participants were used as a covariate in the statistical analysis to explain a part of the variation that would otherwise be treated as the error variance. The result was a 2 (background noise condition) \times 2 (gender) \times 2 (nationality) mixed design MANCOVA with noise sensitivity as a continuous covariate, and both RA and RT as two dependent variables representing cognitive performance in each background noise condition. Pillai's trace has been used as the test statistic for the MANCOVA (Field, 2013). When an interaction between the variables was detected, follow-up analyses were carried out using one-way ANOVA to break down the interaction. Since there were only two background noise conditions, the assumption of sphericity was met and no correction was applied to the reported *F*-ratios from the ANOVA tests.

6.11.2 Univariate approach

The RA score for the information matching task and one-back task showed ceiling effects i.e., near perfect scores in both background noise conditions. In this case, the multivariate approach was not possible as only one dependent variable (RT) was available to be analysed. Therefore, a univariate mixed design ANCOVA was adopted for these two tasks. The result was a 2 (background noise condition) \times 2 (gender) \times 2 (nationality) mixed design ANCOVA with noise sensitivity as a continuous covariate, and RT as the dependent variable representing cognitive performance in each background noise condition.

6.11.3 Checking assumptions

The statistical models used in this study are based on assumptions. The data used to run the statistical models should meet a set of assumptions in order to give accurate results. Visual inspections using Q-Q plots and the Shapiro-Wilk normality test were used to check the assumption of normal distribution for both response accuracy and reaction time of all tasks in both background noise conditions.

Using the Shapiro-Wilk test, the response accuracy score under the speech-only condition was found to significantly deviate from a normal distribution, whilst it did not deviate from a normal distribution under the masked speech condition. Reaction times in both background noise conditions were found to be positively skewed, therefore the common logarithm of the scores was taken to transform the skewed distribution back to normal distribution. After the transformation, the Shapiro-Wilk normality test revealed that reaction times did not deviate significantly from normal. The data were also checked for outliers using box plots and the standard deviation, SD. Scores greater/smaller than the mean ± 2.5 SD were treated as outliers and removed from the analysis.

The deviations from a normal distribution for the response accuracy score under the speech-only condition should not cause concerns, and there are good reasons for that. Firstly, normality tests, such as Shapiro-Wilk, are notorious for being too sensitive to the sample size, i.e., they tend to show statistically significant results in large samples even for small and unimportant effects, while they show statistically insignificant results for large and important effects in smaller samples. Secondly, as the sample size gets larger, the assumption of normality becomes less important, because sampling distribution will be normal regardless of the distribution of the population from which the sample is drawn. This is known as the central limit theorem. Field (2013) suggests that in a light tailed distribution where outliers are rare, a sample size as small as 20 could suffice to assume normality using the central limit theorem. Lastly, the ANOVA tests (which cover variations of MANOVA) are quite robust to the violation of normality when an equal number of participants is used in the experimental conditions. The current experiment was a repeated measure design in which all participants took part in both conditions, which means an equal number of participants in both conditions. Therefore, the deviation from normality in the response accuracy score should not bias the statistical analysis.

The assumption of equality of covariance matrices for the MANCOVA was tested using Box's test (Field, 2013), which tests the null hypothesis that the observed covariance matrices of the dependent variables (i.e., RA and RT) are equal across groups.

The assumption of homogeneity of variances across groups was tested using Levene's test (Field, 2013) which tests the null hypothesis that variances in different groups are equal. This test is only relevant to tests where different participants took part in different conditions, i.e., when testing differences between males and females, and between native and non-native speakers. Attention was paid to have equal numbers of male/female and native/non-native participants to minimise the effect of violation of the assumption of homogeneity of variances, which gives more credibility to the statistical analysis.

6.11.4 Non-parametric tests

Wilcoxon signed-rank (Field, 2013) was used alongside the Bias-corrected and accelerated bootstrap 95% confidence intervals for the statistical analysis of the responses from the satisfaction questionnaire, subjective task difficulty and the subjective speech intelligibility. For this type of data, the assumptions on which parametric tests are based could not be assumed and therefore their non-parametric statistical models were used.

6.11.5 Effect size

There were only two test conditions in this experiment, speech only and masked speech. This situation results in an F -test which has 1 degree of freedom for the model. When there is only 1 degree of freedom, then the F -ratio reported as a part of ANOVA (and MANOVA) models can be converted to the effect size (r) using the following equation (Field, 2013):

$$r = \sqrt{\frac{F(1, df_R)}{F(1, df_R) + df_R}} \quad (6.3)$$

where $F(1, df_R)$ is the F -ratio given as a part of the ANOVA (and MANOVA) test; and df_R is the residual degrees of freedom.

Field (2013) recommends reporting r as an effect size, instead of the partial eta squared (partial η^2) readily given as an output in SPSS, as the former can easily be interpreted.

When nonparametric tests were used, i.e., the Wilcoxon signed-rank test, the effect size was calculated using the following equation (Field, 2013):

$$r = \frac{z}{\sqrt{N}} \quad (6.4)$$

where z is the z -score that SPSS produces; and N is the number of total observations.

6.11.6 Normalisation of scores

Response accuracy scores and reaction times were both normalised to have a value between 0 and 1. The normalisation of the response accuracy was carried out by dividing each participant's score by the maximum score that they could achieve, e.g., 9 in the case of the serial recall task. The normalisation for the reaction time was achieved through dividing each participant's reaction time by the longest reaction time recorded in the two background noise conditions. Since the hypothesis of the experiment was that participants score higher and take less time to respond in the masked speech condition, the reaction time was reversed by subtracting each participant's reaction time from the longest recorded reaction time, before the normalisation was performed. The reversed reaction time makes the interpretation of the statistical analysis easier as an improvement in performance can be identified by an increased level in either response accuracy score or reversed reaction time, or both.

6.12 Results

6.12.1 Serial recall task

Scores from all participants were retained for the statistical analysis in the serial recall task, as no outliers were spotted. Box's Test of Equality of Covariance Matrices was not significant ($p = .953$), which suggests that the observed covariance matrices of the dependent variables were equal across groups. Using Pillai's trace, there was a significant effect of background noise condition on performance, $V = 0.096$, $F(1, 43) = 4.563$, $p = .038$, $r = .310$. Performance was defined by the two dependent variables, RA and RT and their interaction. To break down this interaction, separate univariate ANOVAs were carried out on the dependent variables RA and RT. The ANOVA test revealed nonsignificant effects of background noise condition on RA, $F(1, 43) = 3.100$, $p = .085$, $r = .259$, and RT, $F(1, 43) = 0.347$, $p = .559$, $r = .089$. The mean score of RA and RT in both speech conditions are shown in Figure 6.10. It is clear that both RA and RT are higher in the masked speech condition², yet neither RA nor RT was individually high

² The statistical analysis requires RT to be reversed to give people who took less time to complete the task a higher RT score. This may seem as if the reaction time was longer under the masked speech condition compared to the speech-only condition, but the opposite is true.

enough in the masked speech condition to result in a statistically significant difference. However, when the two variables and their interaction were taken together (i.e., using MANOVA), the difference became statistically significant.

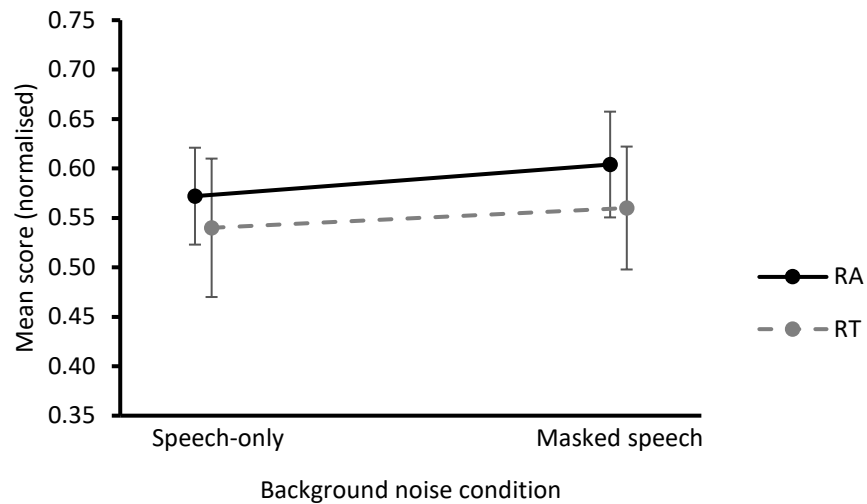


Figure 6.10 Normalised RA and RT (reversed) scores in the serial recall task under the speech-only and the masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

There was a significant main effect of gender on performance, $F(1, 43) = 5.369, p = .025, r = .333$. This result indicates that, generally, performance scores for male and female participants were different irrespective of the background noise conditions. Figure 6.11 shows the averaged performance scores for both male and female participants. The figure suggests that male participants' performance (i.e., the average score of RA and RT in both background noise conditions, corrected for the effect of the covariate NoiseQ) was significantly higher than that of female participants.

There was no interaction between the gender of participants and the background noise condition, $V = 0.036, F(1, 43) = 1.619, p = .210, r = .190$. This result is shown in Figure 6.12, which suggests that the profile of change in performance (i.e., the slope of the lines in Figure 6.12) across the background noise conditions was not statistically different between male and female participants. A closer look at Figure 6.12 reveals that performance of male participants remained unaffected by the background noise condition (shown by a flat line), while the performance of female participants was affected by the background noise condition as they appear to have performed better in the masked speech

condition. However, this increase in performance was not statistically large enough to result in a significant interaction between the gender and background noise conditions.

There was no significant interaction between the gender of participants and the performance score, $V = 0.007$, $F(1, 43) = 0.323$, $p = .573$, $r = .086$. This result is shown in Figure 6.13, and suggests that the difference between RA and RT scores was similar for male and female participants, irrespective of the background noise condition. The result can also be interpreted as the ratio between RA and RT scores being similar for male and female participants.

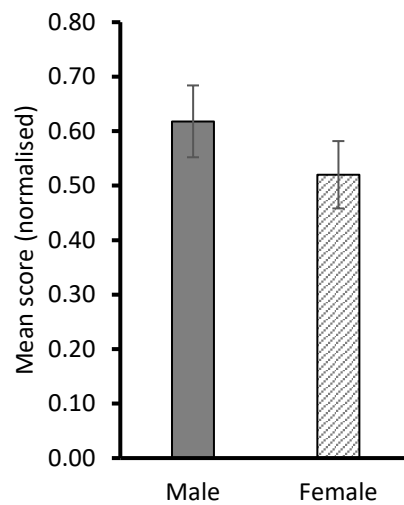


Figure 6.11 Averaged normalised performance score in the serial recall task for male and female participants, irrespective of background noise condition. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245

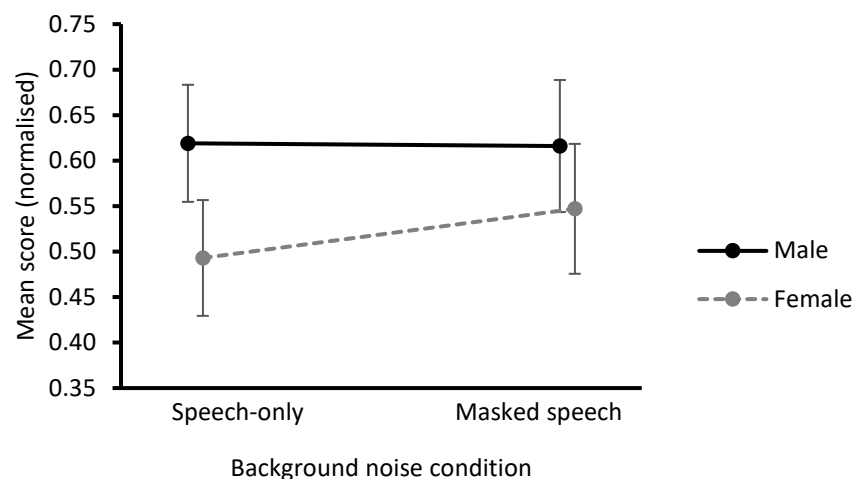


Figure 6.12 Averaged normalised performance score (RA and RT) in the serial recall task for male and female participants under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

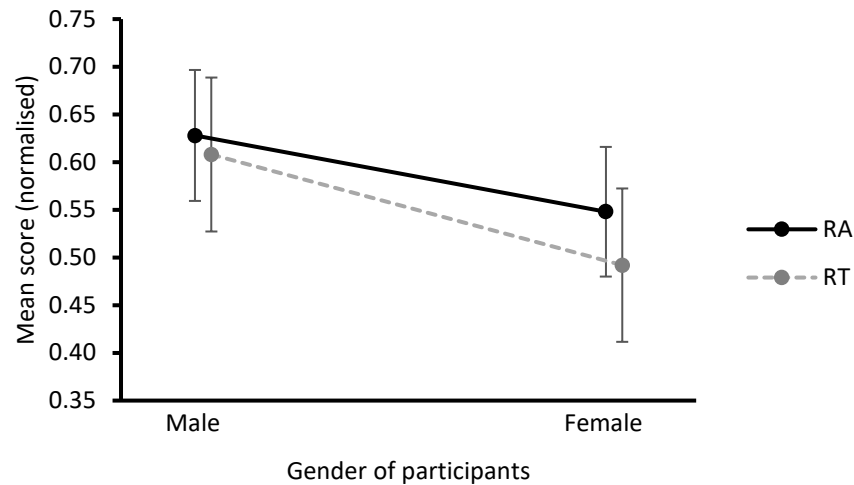


Figure 6.13 Normalised RA and RT (reversed) scores in the serial recall task for male and female participants, irrespective of the background noise condition. Error bars represent 95% CI. Scores are normalised for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

The three-way gender \times background noise condition \times performance interaction was also non-significant, $V = 0.001$, $F(1, 43) = 0.055$, $p = .887$, $r = .036$. This interaction is shown in Figure 6.14. This result indicates that the profile of change in RA and RT scores between the two background noise conditions was similar for male and female participants, which confirms the two-way gender \times background noise condition interaction results.

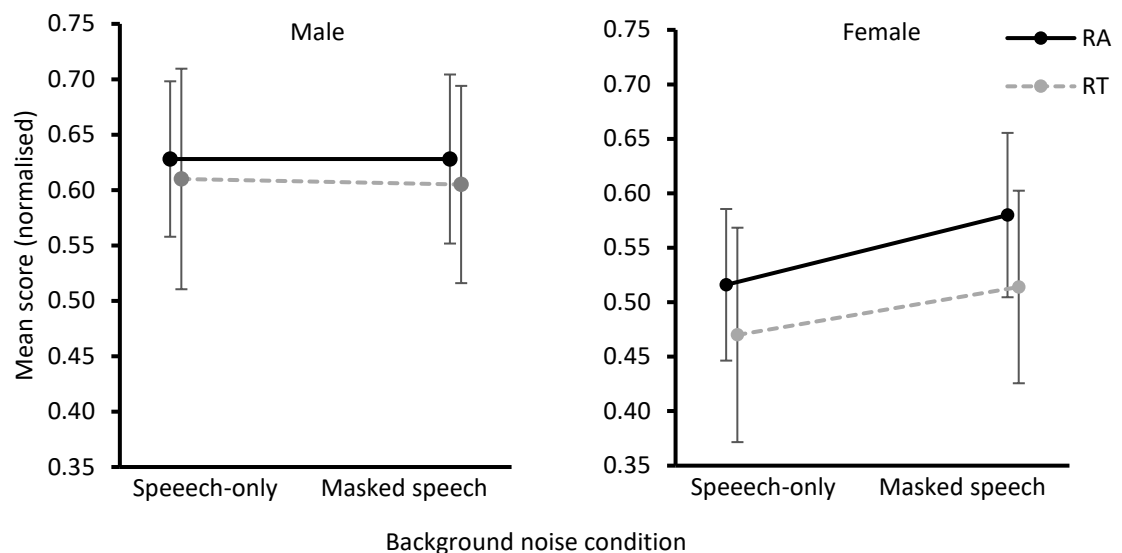


Figure 6.14 Normalised RA and RT (reversed) scores in the serial recall task under speech-only and masked speech conditions for both male and female participants. Error bars represent 95% CI. Scores are normalised for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

It would appear that male participants did not benefit from the masking system as their performance remained almost the same in both conditions, while female participants performed better in the masked speech condition both in terms of reaction time and response accuracy. It is unclear why this has happened, but one possible explanation would be that male participants might have been more capable of ignoring the background noise, hence their steady performance. Female participants, on the other hand, might have been more sensitive to the irrelevant speech and found it difficult to concentrate under the speech-only condition. When the water sound was added to the background noise, the background noise might have become less distracting for female participants, hence the increase in their performance. However, this is not reflected in the self-rated noise sensitivity score, as the average scores for male and female participants were similar. Having said that, the noise sensitivity is about noise generally and not irrelevant speech particularly. The background noise in the experiment was irrelevant speech and female participants' attention might have been drawn to it, which might have negatively affected their performance.

The main effect of nationality (native vs. non-native English speakers) of participants was not significant, $F(1, 43) = 0.332$, $p = .568$, $r = .088$. This effect is shown in Figure 6.15, which suggests that, generally, performance scores were similar for native and non-native English speakers, regardless of the background noise conditions.

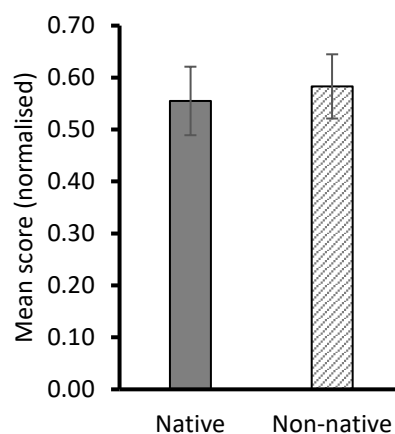


Figure 6.15 Global normalised performance score (RA and RT) in the serial recall task for native and non-native English speakers. Error bars represent 95% CI. Scores are normalised for the covariate NoiseQ = 2.245.

The interaction between background noise conditions and the nationality of participants was not statistically significant, $V = 0.013$, $F(1, 43) = 0.576$, $p = .452$, $r = .115$. This effect indicates that the profile of change in performance from one background noise condition

to the other was similar for native and non-native English speakers. The interaction is shown in Figure 6.16.

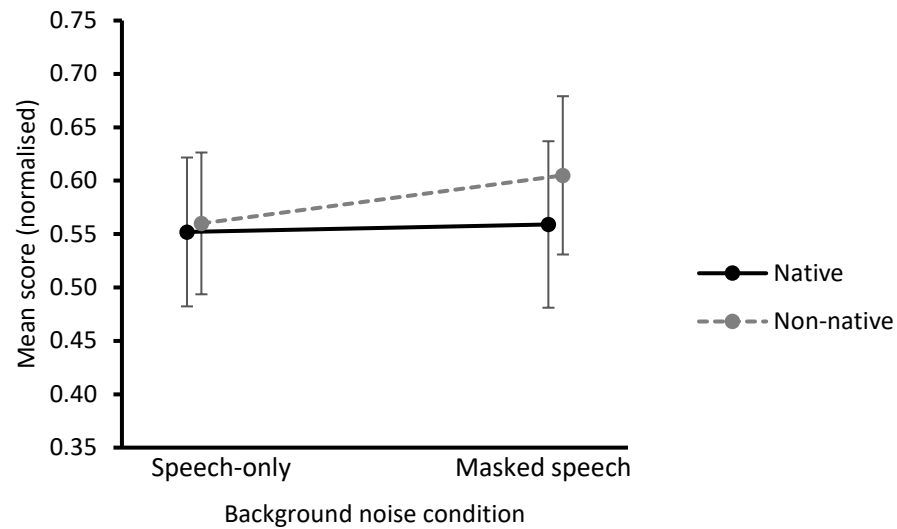


Figure 6.16 Averaged normalised performance score in the serial recall task for native and non-native English speakers under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

There was a significant interaction between nationality of participants and performance (average of RA and RT), $V = 0.153$, $F(1, 43) = 7.757$, $p = .008$, $r = .391$. This interaction is shown in Figure 6.17, which suggests that the difference between RA and RT scores is significantly different between native and non-native participants.

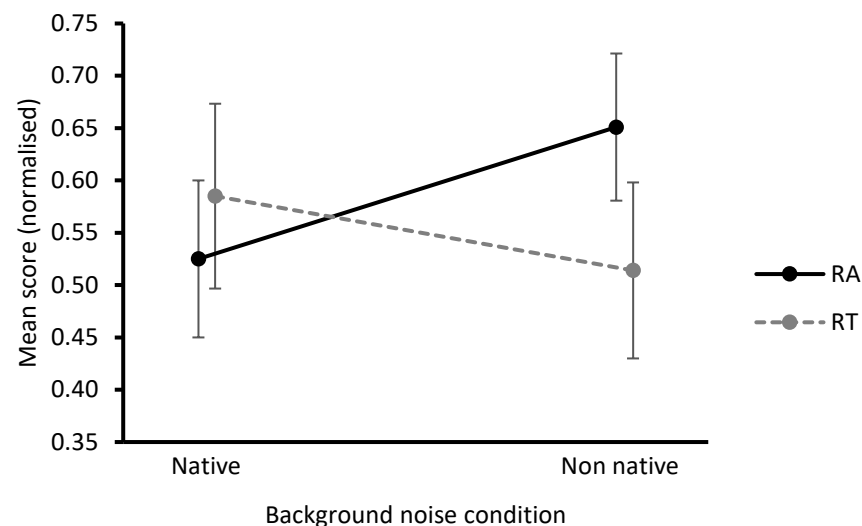


Figure 6.17 Normalised RA and RT (reversed) scores in the serial recall task for native and non-native English speakers irrespective of the background noise condition. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

As Figure 6.17 shows, there is a different trade-off between RA and RT for native and non-native speakers. This suggests that ignoring the background noise conditions, non-native participants had a higher response accuracy but took longer to complete, in comparison to the native participants, who had a lower response accuracy but were quicker to complete the task. Performing further comparisons on this interaction revealed that the non-native participants' RA was significantly higher than the RA of native participants, $F(1, 43) = 5.390, p = .025, r = .334$, yet, the non-native participants' RT was not significantly longer, $F(1, 43) = 1.220, p = .275, r = .166$. It is interesting that the main effect of nationality on performance reported earlier revealed a nonsignificant difference in performance between native and non-native participants when performance was defined by a combined effect of both RA and RT. However, when performance is defined by only RA and RT is ignored, the difference in performance between native and non-native participants becomes statistically significant. This highlights the importance of considering both time and response accuracy as two variables defining cognitive performance. This also partially stands for the contradicting results reported by different studies.

The three-way nationality \times background noise condition \times performance interaction was non-significant, $V < 0.001, F(1, 43) = 0.020, p = .887, r = .022$. This interaction is shown in Figure 6.18. This result indicates that the profile of change in RA and RT scores across the two background noise conditions was similar for native and non-native participants. Native participants showed a marginal increase in RA and RT in the masked speech condition. Non-native participants, on the other hand, showed a slightly higher increase in RA and RT in the masked speech condition compared to native speakers, but not enough to result in a statistically significant interaction.

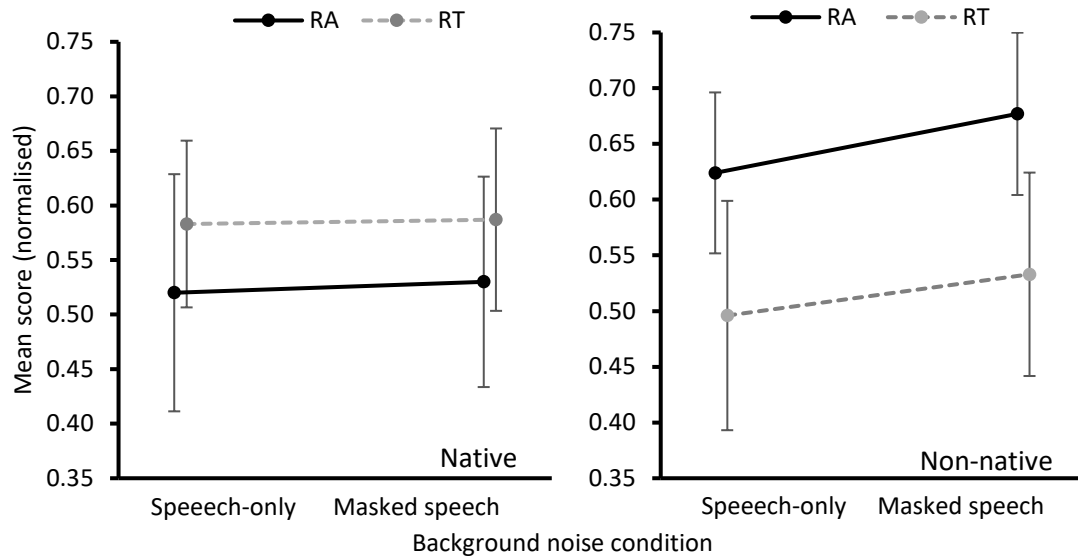


Figure 6.18 Normalised RA and RT (reversed) scores in the serial recall task for native and non-native English speakers under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

The main effect of the covariate NoiseQ was not statistically significant, $F(1, 43) = 1.858$, $p = .180$, $r = .204$, which indicates that the covariate did not significantly predict the change in dependent variables (i.e., RA and RT) across the two background noise conditions. However, the effect size ($r = .204$) suggests that despite the non-significant figure, NoiseQ accounted for a portion of the variation that would have otherwise been considered as error in the statistical model.

There was also no significant interactions between NoiseQ and the background noise condition, $V = 0.014$, $F(1, 43) = 0.623$, $p = .434$, $r = .111$, nor between NoiseQ and performance, $V = 0.003$, $F(1, 43) = 0.111$, $p = .954$, $r = .051$. However, there was a significant 3-way NoiseQ \times background noise condition \times performance interaction, $V = 0.106$, $F(1, 43) = 5.094$, $p = .029$, $r = .325$. This 3-way interaction indicates that the participants' RA and RT scores under the speech-only and masked speech conditions were differently affected by their sensitivity to noise (i.e., NoiseQ score). This interaction is illustrated in Figure 6.19. The figure suggests that under the speech-only condition, participants' RA was not affected by their sensitivity to noise, but their RT was positively correlated with NoiseQ i.e., the more sensitive the participants were, the longer they took to complete the task. Under the masked speech condition, both RT and RA had a positive correlation with participants' NoiseQ score, suggesting that the participants who were more sensitive to noise scored a higher response accuracy, but took longer to complete

the task. These interactions are further analysed in Section 6.12.7, with correlation analysis, to estimate the effect size of the correlations and their significance.

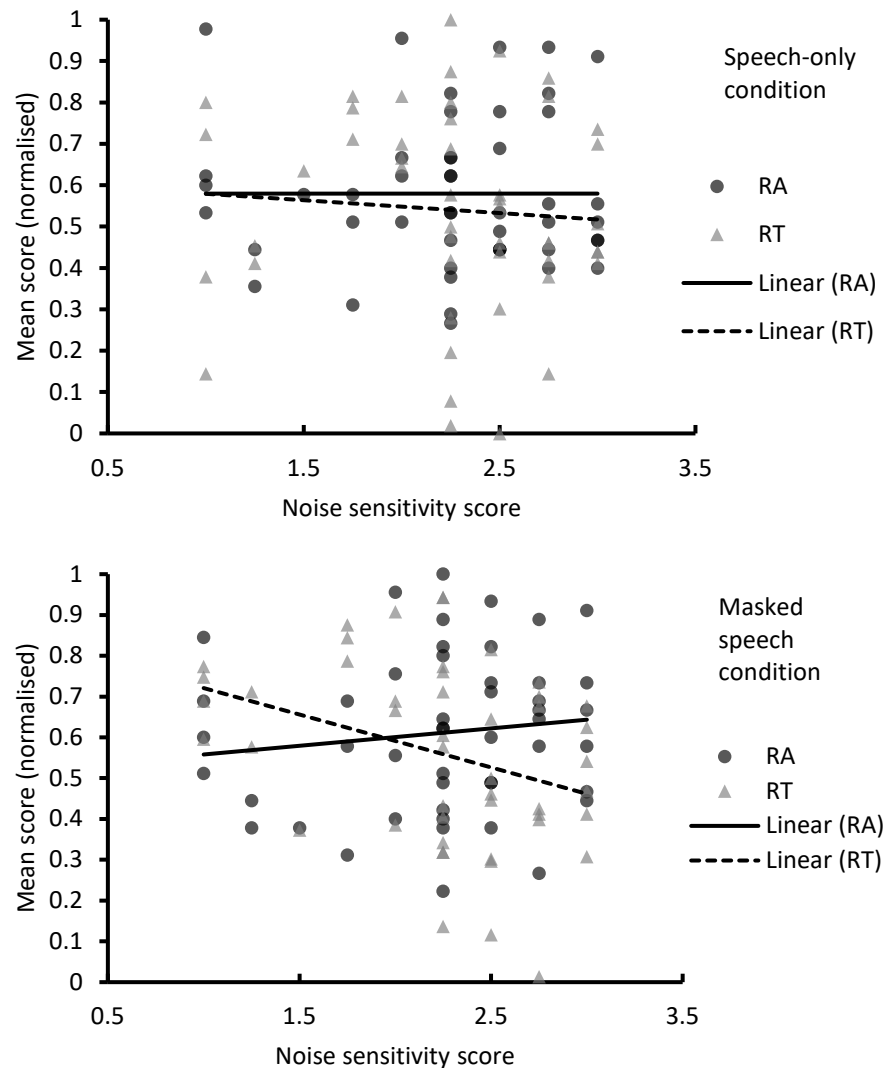


Figure 6.19 Relationship between normalised RA and RT (reversed) scores in the serial recall task with NoiseQ scores under speech-only and masked speech conditions. Linear regression was used to generate the trend lines.

Finally, there was no statistically significant interaction between gender \times nationality, $F(1, 43) = 1.534$, $p = .222$, $r = .186$, nor between gender \times nationality \times background noise condition, $V = 0.005$, $F(1, 43) = 0.200$, $p = .657$, $r = .068$, nor between gender \times nationality \times performance, $V = 0.027$, $F(1, 43) = 0.581$, $p = .450$, $r = .115$.

6.12.2 Information matching task

Box's Test of Equality of Covariance Matrices was not significant ($p = .668$), which suggests that the observed covariance matrices of the dependent variables were equal across groups. As mentioned earlier, the RA for this task had shown a ceiling effect, and

therefore, the statistical analysis is based on RT only. Two participants (1 male, 1 female, both non-native English speakers) had taken too long to complete the task (score > mean + 2.5 SD) and therefore, their RT scores were treated as outliers and removed from the statistical analysis (i.e., $N = 46$).

The main effect of background noise condition on RT was not statistically significant, $F(1, 41) = 1.050$, $p = .311$, $r = .158$. The mean RT scores in both background noise conditions are shown in Figure 6.20. It appears that the water sound had no effect on the performance of participants in the information matching task, as their RT scores are almost identical in both background noise conditions.

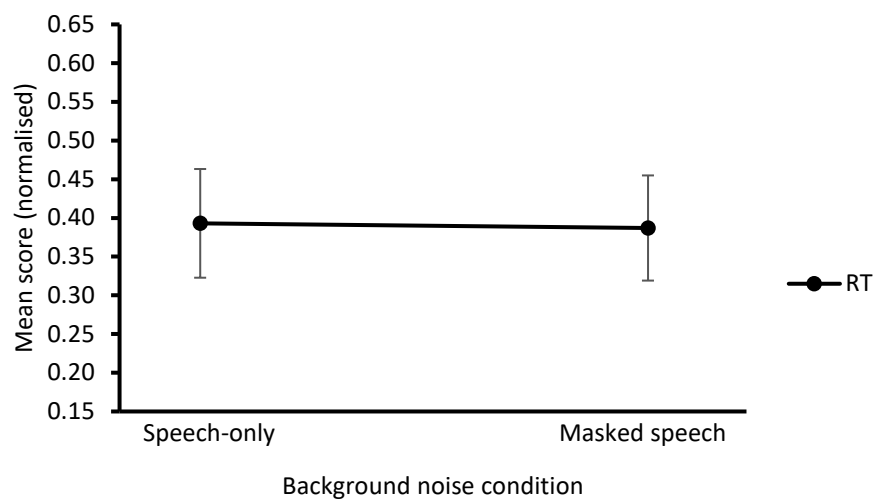


Figure 6.20 Normalised RT score (reversed) in the information masking task, under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245.

The main effect of gender was not statistically significant, $F(1, 41) = 0.148$, $p = .703$, $r = .060$. The mean RT scores for males and females are shown in Figure 6.21. This result suggests that males' and females' reaction times were similar, irrespective of the background noise condition. There was no significant interaction between gender and background noise conditions, $F(1, 41) = 0.356$, $p = .554$, $r = .093$. This result is shown in Figure 6.22, which indicates that the profile of change in RT across the background noise conditions (i.e., the slope of the lines in Figure 6.22) was not statistically different between male and female participants.

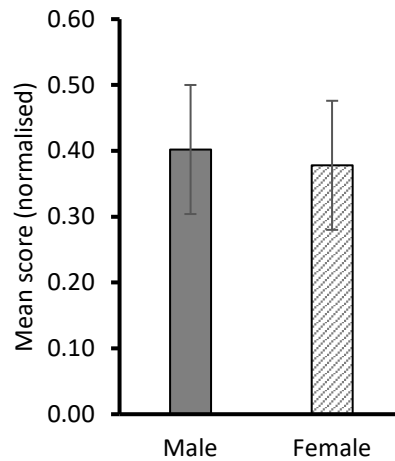


Figure 6.21 Averaged normalised RT score (reversed) in the information masking task for male and female participants. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245.

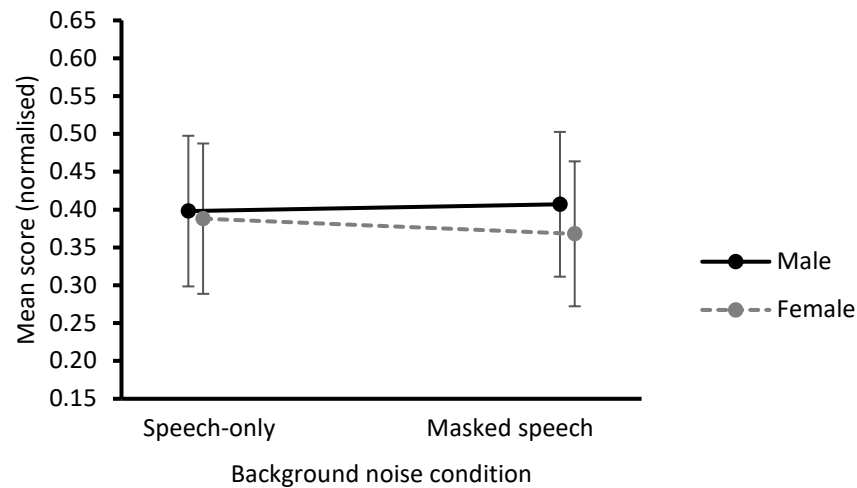


Figure 6.22 Averaged normalised RT score (reversed) in the information masking task for male and female participants under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

The main effect of nationality of participants was statistically significant, $F(1, 41) = 6.124$, $p = .018$, $r = .360$. This effect is shown in Figure 6.23. The figure suggests that native participants' reaction time was significantly shorter than the non-native participants, irrespective of the background noise condition. A similar pattern was detected in the serial recall task as non-native participants took longer to complete the task but they had a higher response accuracy score.

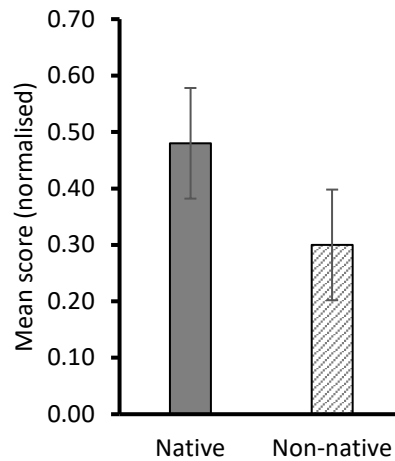


Figure 6.23 Averaged normalised RT score (reversed) in the information masking task for native and non-native participants. Error bars represent 95% CI. scores are corrected for the covariate NoiseQ = 2.245.

There was no interaction between nationality of participants and background noise conditions, $F(1, 41) = 1.834$, $p = .183$, $r = .207$. This result is shown in Figure 6.24 and suggests that the profile of change in RT across the background noise conditions (i.e., the slope of the lines in Figure 6.24) was not statistically different between native and non-native participants. The figure shows that native participants took slightly longer to complete the task under the masked speech condition, while non-native participants were slightly faster under the masked speech condition. However, these differences were not strong enough to result in a statistically significant interaction.

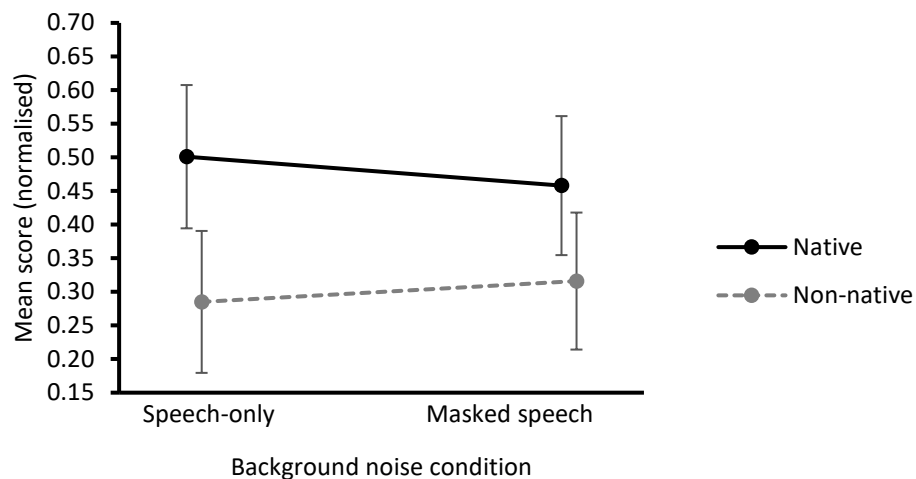


Figure 6.24 Averaged normalised RT score (reversed) in the information masking task for native and non-native participants under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245.

Lines are slightly shifted to avoid overlapped error bars.

The main effect of the covariate NoiseQ was not statistically significant, $F(1, 41) = 0.380$, $p = .541$, $r = .096$, which indicates that the covariate did not significantly predict the change in the dependent variable (i.e., RT) across the two background noise conditions. In addition, there was no significant interaction between NoiseQ and the background noise condition, $F(1, 41) = 1.834$, $p = .276$, $r = .207$. This result is illustrated in Figure 6.25 and shows that participants' reaction time was similarly affected by their sensitivity to noise in both background noise conditions. The graph suggests that the more sensitive participants were, the more time they tended to take to complete the task. The magnitude of this relationship is determined through correlation analysis in Section 6.12.7.

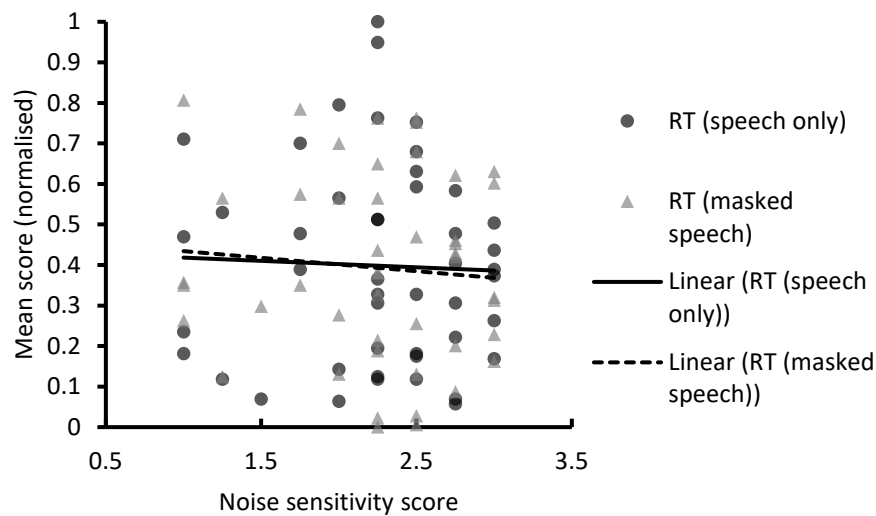


Figure 6.25 Relationship between normalised RT scores (reversed) in the information masking task with NoiseQ score under speech-only and masked speech conditions. Linear regression was used to generate the trend lines.

Finally, there was no statistically significant interactions between gender \times nationality, $F(1, 41) = 1.806$, $p = .186$, $r = .205$, nor between gender \times nationality \times background noise condition, $F(1, 41) = 0.075$, $p = .786$, $r = .043$.

6.12.3 One-back task

Box's Test of Equality of Covariance Matrices was not significant ($p = .551$), which suggests that the observed covariance matrices of the dependent variables were equal across groups. As mentioned earlier, the RA for this task had shown a ceiling effect, and therefore the statistical analysis is based on RT only.

The main effect of background noise conditions on RT was not statistically significant, $F(1, 43) = 1.726$, $p = .196$, $r = .196$. The mean RT scores in both background noise conditions are shown in Figure 6.26. It appears that the water sound had a positive impact

on the reaction time as participants took less time completing the task in the masked speech condition. However, this improvement was not statistically significant.

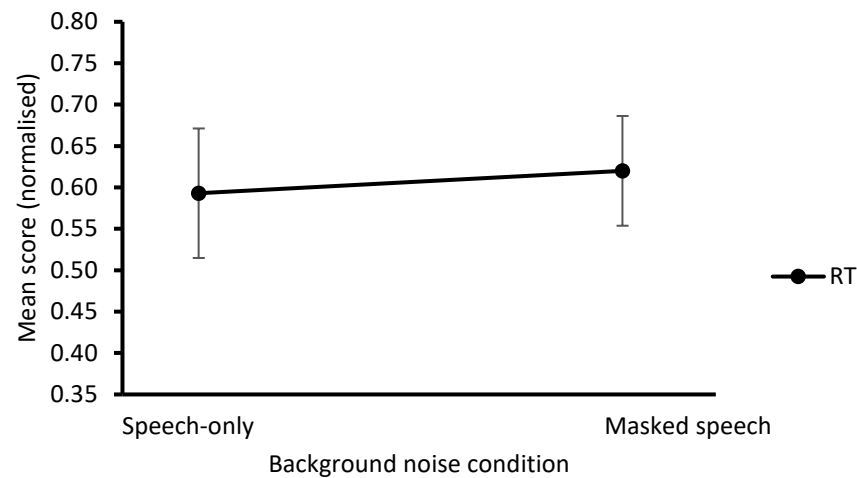


Figure 6.26 Normalised RT scores (reversed) in the one-back task under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245.

The main effect of gender was statistically significant, $F(1, 43) = 5.346$, $p = .026$, $r = .332$. The mean RT scores for males and females are shown in Figure 6.27. This result suggests that males' reaction time was significantly shorter than that of their female counterparts. There was no significant interaction between gender and background noise conditions, $F(1, 43) = 0.656$, $p = .423$, $r = .123$. This result is shown in Figure 6.28 and it indicates that the profile of change in RT from one background noise condition to the other (i.e., the slope of the lines in Figure 6.28) was not statistically different between male and female participants. It appears that male participants did not benefit from the water sound as their RTs are very similar in both speech-only and masked speech conditions. Female participants, on the other hand, benefited from the water sound as their reaction time was shorter in the masked speech condition, compared to the speech-only condition. A similar trend was detected in the serial recall task, where female participants benefited more from the water sound. It is not clear why this happened, also considering that the noise sensitivities of males and females were comparable. One explanation could be that female participants were more easily disturbed by the background noise, so the effect of irrelevant speech was greater on them.

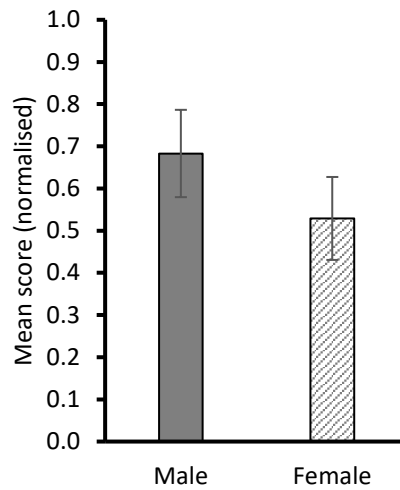


Figure 6.27 Averaged normalised RT score (reversed) for male and female participants in the one-back task. Error bars represent 95% CI. scores are corrected for the covariate NoiseQ = 2.245.

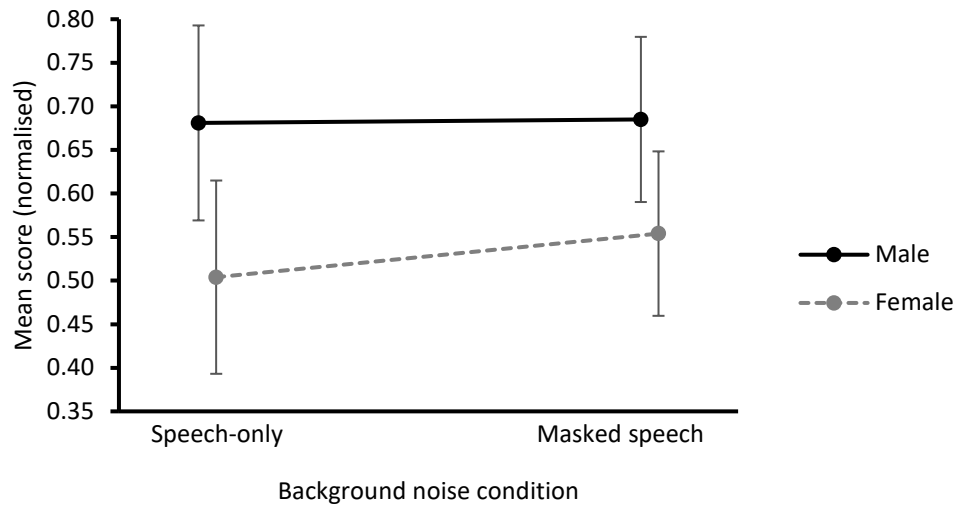


Figure 6.28 Averaged normalised RT score (reversed) for male and female participants in the one-back task under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

The main effect of nationality of participants was not statistically significant, $F(1, 43) = 0.203$, $p = .655$, $r = .069$. This effect is shown in Figure 6.29. The figure indicates that RT scores for both native and non-native participants were similar, irrespective of the background noise condition. Native participants took a slightly shorter time to complete the task, but this might have happened solely by chance.

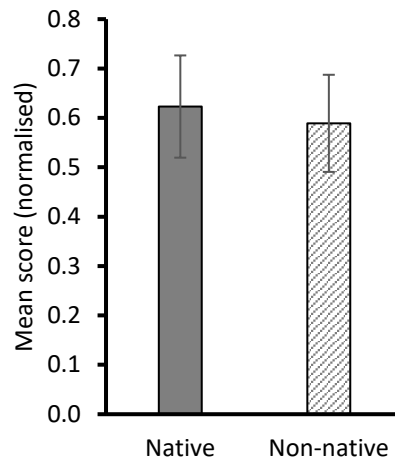


Figure 6.29 Averaged normalised RT score (reversed) for native and non-native participants in the one-back task. Error bars represent 95% CI. scores are corrected for the covariate NoiseQ = 2.245.

There was no interaction between nationality and background noise conditions, $F(1, 43) = 0.652$, $p = .424$, $r = .122$. This result is shown in Figure 6.30 and suggests that the profile of change in RT from one background noise condition to the other (i.e., the slope of the lines in Figure 6.30) was not statistically different between native and non-native participants. It appears that RT scores for both native and non-native participants were similar under the speech-only condition, while under the masked speech condition, native participants' RT (reversed) score was higher i.e., they took less time to complete the test. However, this interaction was not statistically significant as mentioned above.

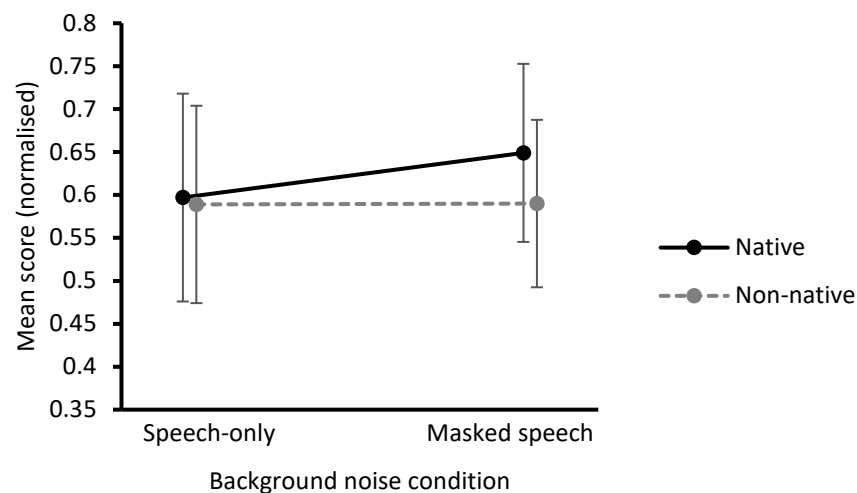


Figure 6.30 Averaged normalised RT score (reversed) for native and non-native participants in the one-back task, under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

The main effect of the covariate NoiseQ was not statistically significant, $F(1, 43) = 0.073$, $p = .788$, $r = .041$, which indicates that the covariate did not significantly predict the change in the dependent variable (i.e., RT) across the two background noise conditions. In addition, there was no significant interaction between NoiseQ and the background noise conditions, $F(1, 43) = 1.269$, $p = .266$, $r = .169$. This result is illustrated in Figure 6.31.

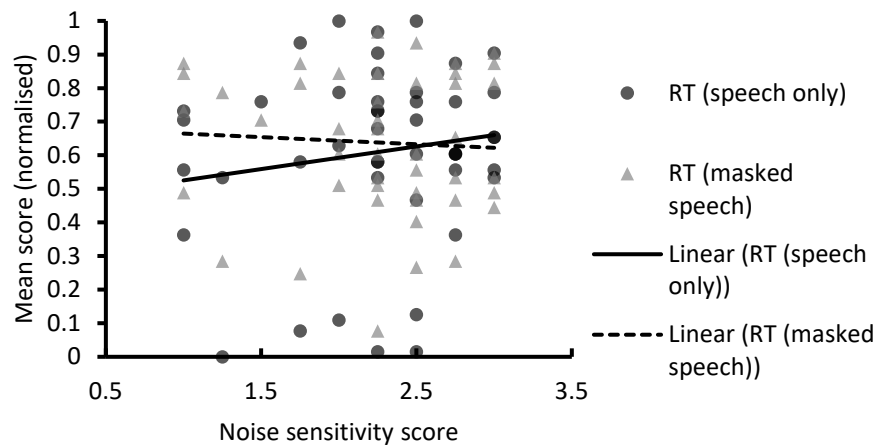


Figure 6.31 Relationship between normalised RT scores (reversed) with NoiseQ score in the one-back task under speech-only and masked speech conditions. Linear regression was used to generate the trend lines.

The figure shows that under the speech-only condition, reaction time reduced with increasing the NoiseQ score, while the trend was reversed in the masked speech condition, as reaction time increased with increasing the NoiseQ score. Since the statistical analyses were not significant, these trends cannot be trusted as they might have happened solely by chance. The magnitude of this relationship is further examined through correlation analysis in Section 6.12.7.

Finally, there was no statistically significant interaction between gender \times nationality, $F(1, 43) < 0.001$, $p = .985$, $r = .004$, nor between gender \times nationality \times background noise condition, $F(1, 43) = 0.321$, $p = .574$, $r = .086$.

6.12.4 Reading comprehension task

Two participants (1 male, 1 female, both non-native English speakers) had taken too long to complete the task (RT scores $>$ mean score + 2.5 SD). Therefore, their scores were treated as outliers and excluded from the statistical analysis (i.e., $N = 46$). Box's Test of Equality of Covariance Matrices was not significant ($p = .075$), which suggests that the observed covariance matrices of the dependent variables were equal across groups.

There was no significant effect of background noise condition on performance, $V = 0.012$, $F(1, 41) = 0.482$, $p = .492$, $r = .108$. This result is presented in Figure 6.32, which shows that neither RA nor RT score was significantly affected by the background noise condition.

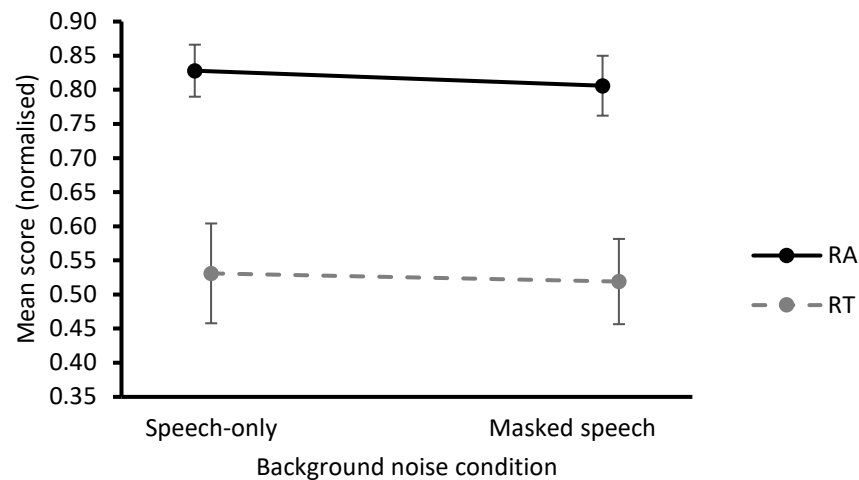


Figure 6.32 Normalised RA and RT (reversed) scores in the reading comprehension task, under speech-only condition and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

There was no significant main effect of gender on performance, $F(1, 41) = 1.826$, $p = .184$, $r = .206$. This result indicates that ignoring the background noise condition, performance scores for male and female participants were similar. This main effect is shown in Figure 6.33. The figure shows that male participants' performance (i.e., the average score of RA and RT in both background noise conditions, corrected for the covariate NoiseQ) was higher than that of female participants, but not high enough to result in a statistically significant difference.

There was no interaction between the gender of participants and the background noise condition, $V = 0.001$, $F(1, 41) = 0.048$, $p = .828$, $r = .034$. This result is shown in Figure 6.34, which suggests that the profile of change in performance (i.e., the slope of the lines in Figure 6.34) across the background noise conditions was not statistically different between male and female participants.

There was no significant interaction between the gender of participants and the performance score, $V < 0.001$, $F(1, 41) = 0.002$, $p = .968$, $r = .007$. Figure 6.35 shows this result, which suggests that the difference between RA and RT is similar for male and

female participants, irrespective of the background noise condition. The result can also be interpreted as the ratio between RA and RT being the same for male and female participants.

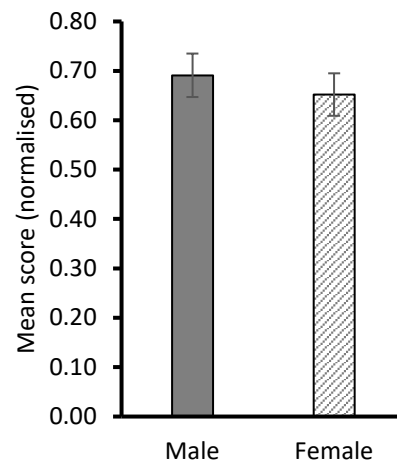


Figure 6.33 Averaged normalised performance score for male and female participants in the reading comprehension task. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245.

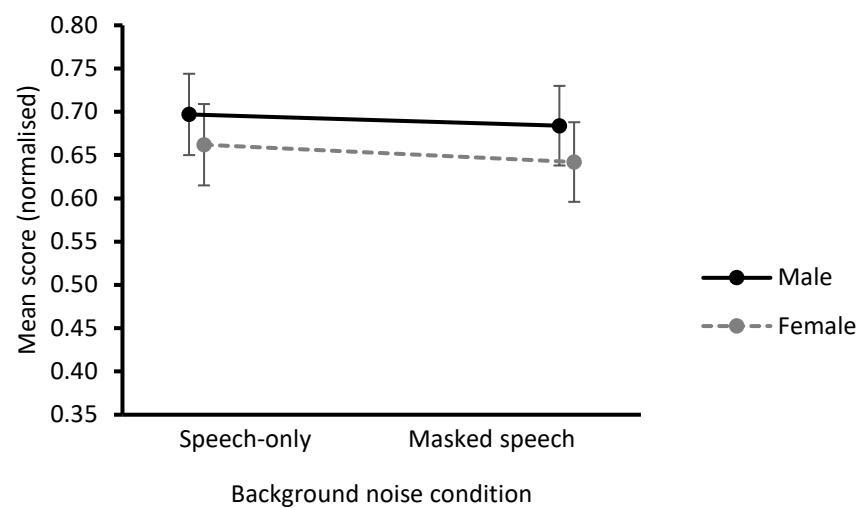


Figure 6.34 Averaged normalised performance score for male and female participants in the reading comprehension task under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

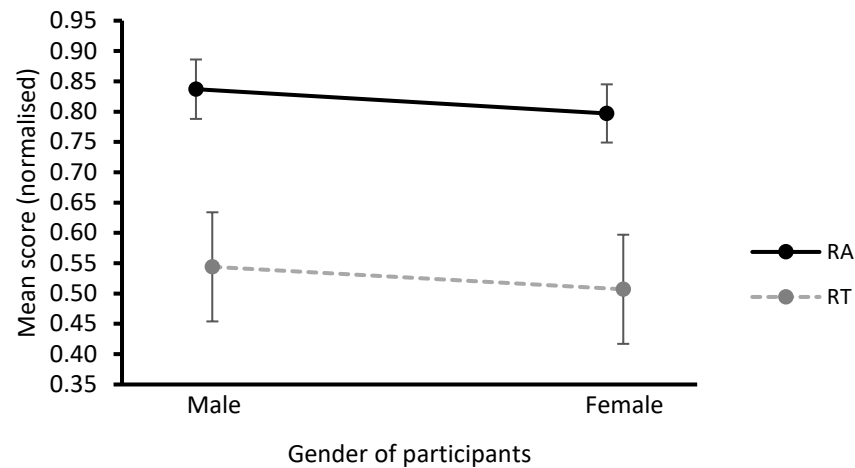


Figure 6.35 Normalised RA and RT (reversed) scores for male and female participants in reading comprehension task, irrespective of background noise condition. Error bars represent 95% CI. Scores are normalised for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

The three-way gender \times background noise condition \times performance interaction was also non-significant, $V < 0.001$, $F(1, 41) = 0.007$, $p = .935$, $r = .013$. This result is shown in Figure 6.36, which indicates that the profile of change in RA and RT scores across the two background noise conditions was similar for male and female participants, which confirms the two-way gender \times background noise condition interaction results.

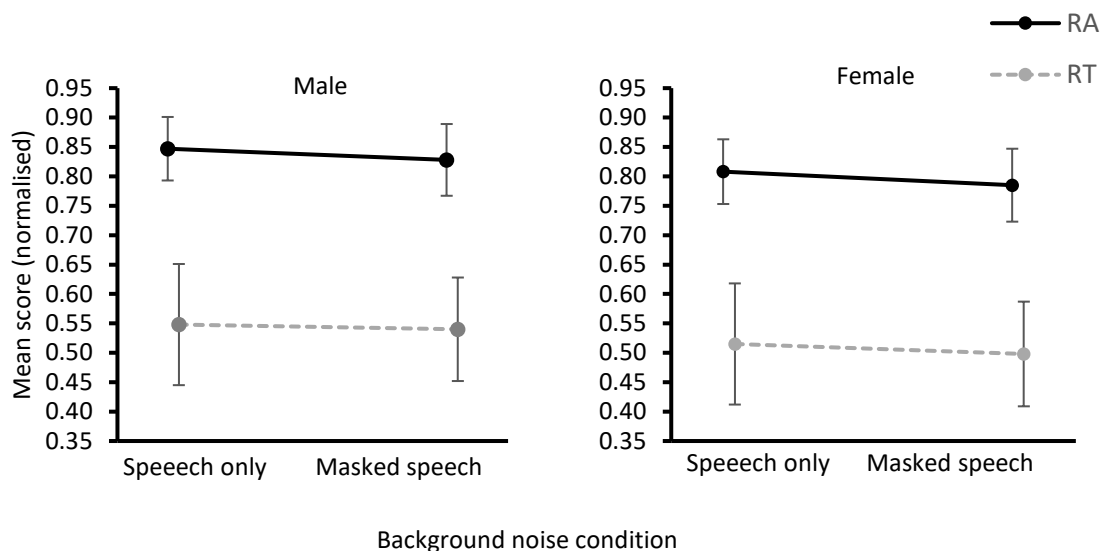


Figure 6.36 Normalised RA and RT (reversed) scores in the reading comprehension task, under speech-only and masked speech conditions for both male and female participants. Error bars represent 95% CI. Scores are normalised for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

There was a significant main effect of nationality of participants on their cognitive performance, $F(1, 41) = 7.091, p = .011, r = .384$. This effect is shown in Figure 6.37, which suggests that cognitive performance of native participants in the reading comprehension task was significantly higher than that of non-native participants, irrespective of the background noise conditions.

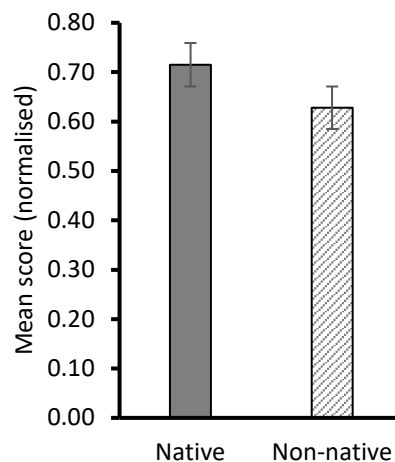


Figure 6.37 Global normalised performance score for native and non-native English speakers in the reading comprehension task. Error bars represent 95% CI. Scores are normalised for the covariate NoiseQ = 2.245.

The interaction between background noise conditions and the nationality of participants was not statistically significant, $V = 0.058, F(1, 41) = 2.519, p = .120, r = .241$. This effect indicates that the profile of change in performance across the two background noise conditions was similar for native and non-native English speakers. The interaction is shown in Figure 6.38. The figure shows a slight increase in performance of native participants under the masked speech condition, whilst there is a decline in performance of non-native participants under the masked speech condition. However, these differences were not strong enough to result in a statistically significant interaction.

There was no significant interactions between the nationality of participants and performance, $V = 0.049, F(1, 41) = 2.124, p = .153, r = .222$. This result is shown in Figure 6.39, which suggests that the difference between RA and RT scores, irrespective of the background noise condition, was similar for native and non-native participants. Although this interaction is not statistically significant, it reveals the reason which most likely resulted in the main effect of nationality being statistically significant. Native and

non-native participants had a similar response accuracy, but the time taken by non-native participants to read the paragraphs and respond was much longer. The linguistic capability might have played a role in causing this difference. Non-native participants might have found it difficult to read the texts, as they were not using their native language. However, once they read and understood the text, their response accuracy was as good as that of native speakers.

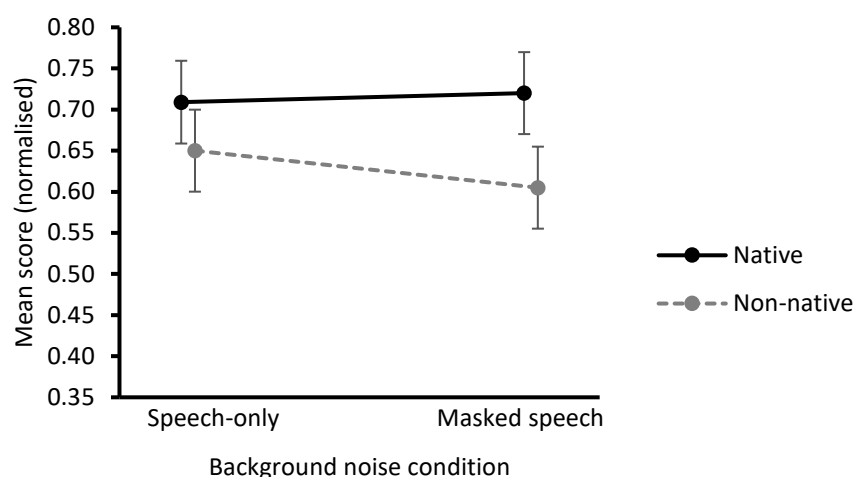


Figure 6.38 Averaged normalised performance scores for native and non-native English speakers in the reading comprehension task under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

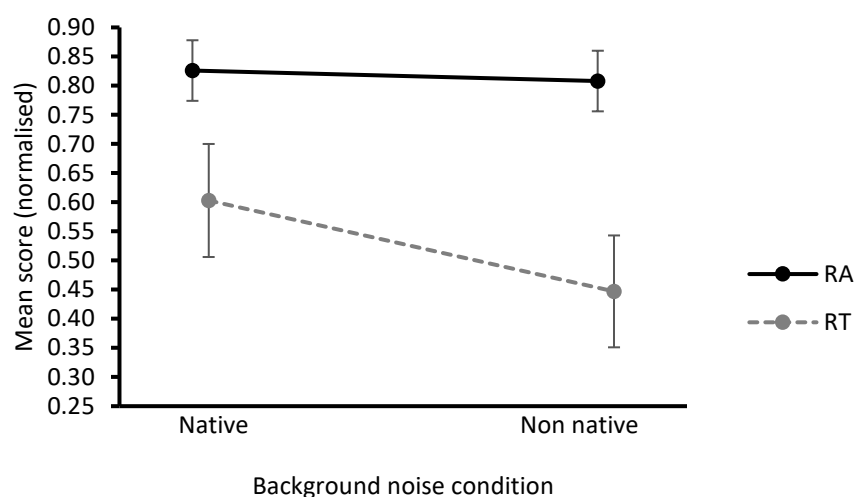


Figure 6.39 Normalised RA and RT (reversed) scores for native and non-native English speakers in the reading comprehension task, irrespective of the background noise condition. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

The three-way nationality \times background noise condition \times performance interaction was not statistically significant, $V = 0.009$, $F(1, 41) = 0.372$, $p = .545$, $r = .095$. This result is shown in Figure 6.40 and indicates that the profile of change in RA and RT scores across the two background noise conditions was similar for native and non-native participants. Native participants showed a marginal increase in RA and RT in the masked speech condition. Non-native participants, on the other hand, showed a slight decline in RA and RT scores in the masked speech condition, but not enough to result in a statistically significant interaction.

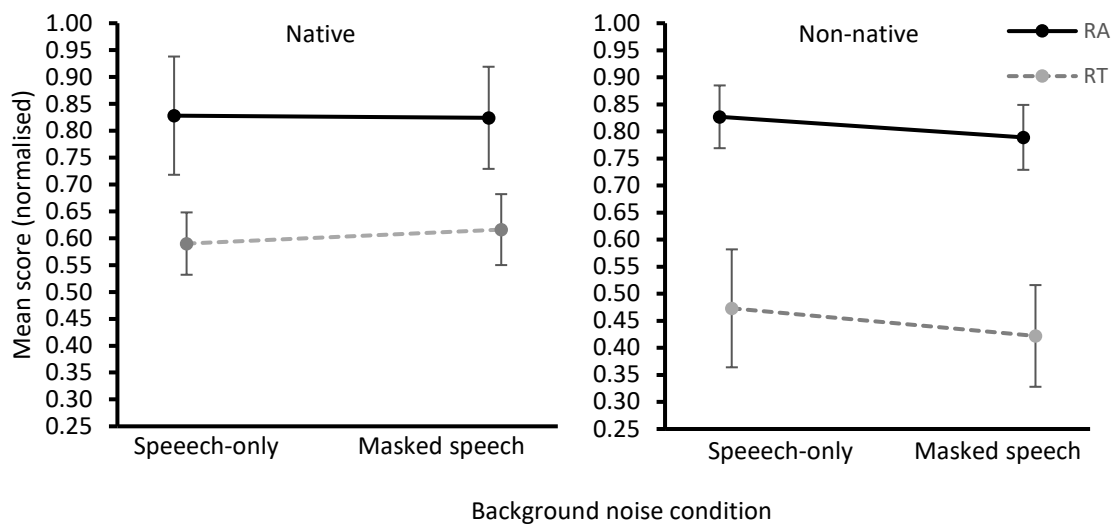


Figure 6.40 Normalised RA and RT (reversed) scores for native and non-native English speakers in the reading comprehension task under speech-only and masked speech conditions. Error bars represent 95% CI. Scores are corrected for the covariate NoiseQ = 2.245. Lines are slightly shifted to avoid overlapped error bars.

The main effect of the covariate NoiseQ was not statistically significant, $F(1, 41) = 0.337$, $p = .565$, $r = .090$, which indicates that the covariate did not significantly predict the change in dependent variables (i.e., RA and RT) across the two background noise conditions.

There was also no significant interaction between NoiseQ and background noise conditions, $V = 0.002$, $F(1, 41) = 0.063$, $p = .803$, $r = .039$, nor between NoiseQ and performance, $V = 0.072$, $F(1, 41) = 3.204$, $p = .081$, $r = .269$. Furthermore, the 3-way interaction NoiseQ \times background noise condition \times performance, was not statistically significant, $V = 0.010$, $F(1, 41) = 0.419$, $p = .521$, $r = .101$. This 3-way interaction, as shown in Figure 6.41, indicates that response accuracy slightly increased with participants' noise sensitivity (i.e., NoiseQ score) in both speech-only and masked speech conditions. Furthermore, the reaction time increased with sensitivity to noise in both

background noise conditions, i.e., the more sensitive to noise the participants were, the more likely they were to take a longer period to complete the task. Correlation analysis is given in Section 6.12.7 to examine whether these relationships are statistically significant.

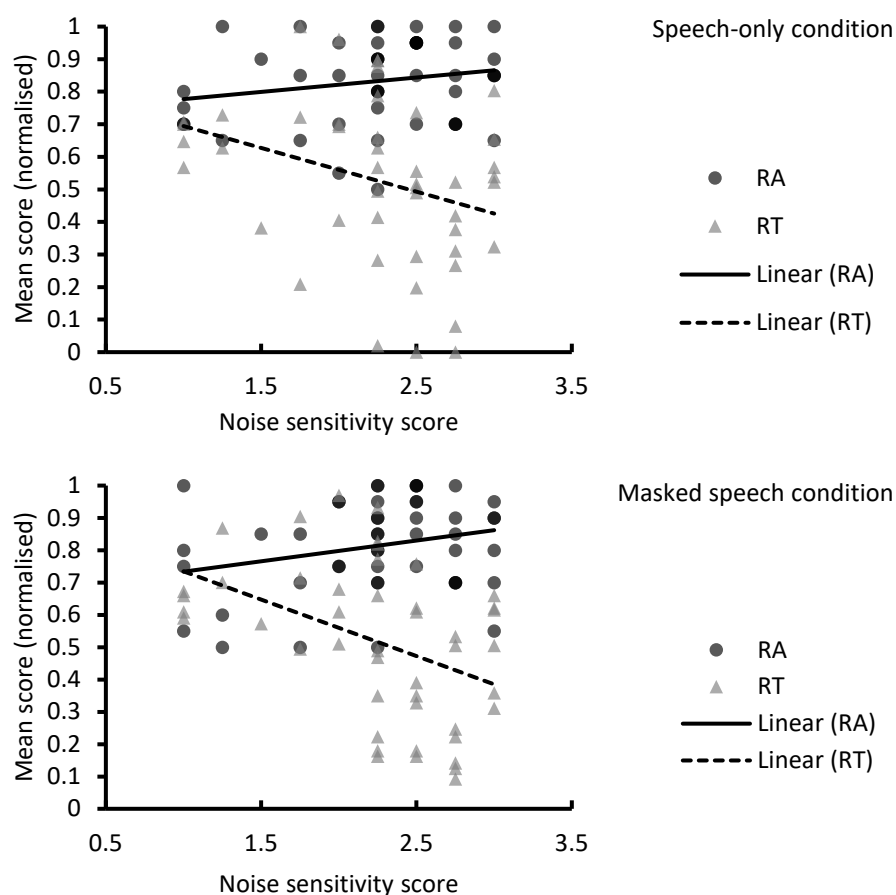


Figure 6.41 Relationship between normalised RA and RT (reversed) scores in the reading comprehension task with NoiseQ score under speech-only and masked speech conditions. Linear regression was used to generate the trend lines.

Finally, there was no statistically significant interactions between gender \times nationality, $F(1, 41) = 1.027$, $p = .317$, $r = .156$, nor between gender \times nationality \times background noise condition, $V = 0.053$, $F(1, 41) = 2.307$, $p = .136$, $r = .231$, nor between gender \times nationality \times performance, $V = 0.004$, $F(1, 41) = 0.181$, $p = .673$, $r = .066$.

The reading comprehension task included four components each measuring a different cognitive ability. The components were *text memory*, *text inferencing*, *knowledge integration* (with 3 levels of difficulty), and *knowledge access* (with two levels of difficulty). Only the overall RT was measured, as the RT for completing questions in each component was not measured. Therefore, the multivariate approach used above would not be applicable to the statistical analysis on the basis of the individual components. A different approach would have been ignoring the reaction time and treating each

component as a dependent variable representing performance. Hence, there would be 4 dependent variables whose combined scores would represent performance in the reading comprehension task. However, this approach was excluded for brevity as there was no sign that the individual components were differently affected by the background noise conditions. Figure 6.42 shows the average scores for each component in the two background noise conditions.

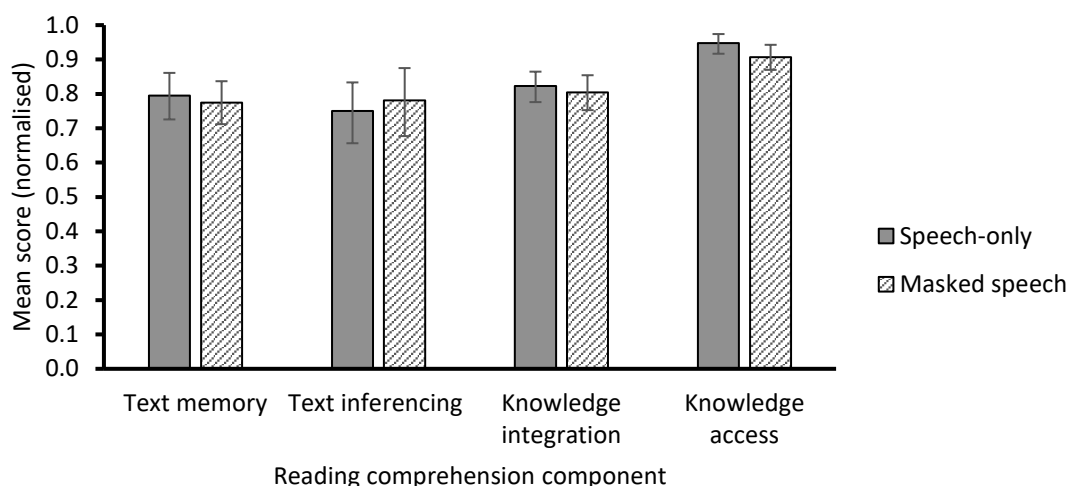


Figure 6.42 Normalised RA score for the reading comprehension components under speech-only and masked speech conditions. Error bars represent 95% CI.

The scores are very similar between the two background noise conditions for each component, and the confidence intervals greatly overlap, which suggest a non-significant difference. Statistically, applying multiple statistical models to the same set of data would result in an inflated familywise error (Type I error) which would increase the risk of finding statistically significant differences which might have happened by chance alone. Therefore, no further statistical analysis was carried out on the data and it was assumed that the statistical analysis made using the overall scores was representative of the individual components.

6.12.5 Subjective satisfaction

The 8 items used in the satisfaction questionnaires showed a good internal consistency (i.e., Cronbach's $\alpha > .7$), in both the speech-only condition, Cronbach's $\alpha = .787$, and the masked speech condition, Cronbach's $\alpha = .773$. These scores suggest that the items included in the satisfaction questionnaire reflected the construct that they were measuring, i.e., the sound environment.

Bias-corrected and accelerated Bootstrap 95% confidence intervals are reported in square brackets. The overall satisfaction score in the speech-only condition was very low, $M = 2.719$ [2.32, 3.132], and increased significantly in the masked speech condition, $M = 4.253$ [3.820, 4.684], $z = 5.537$, $p < .001$, $r = .565$. On the individual items' level, subjective satisfaction was significantly higher in the masked speech condition for all eight items included in the questionnaire, even after adjusting the p -values using the Benjamini-Hochberg procedure to control the false discovery rate. These results are shown in Figure 6.43, where it can be seen that the water sound made the sound environment more pleasant (Item 1), less attention drawing (Item 2), helped people to concentrate on their tasks (Item 3) and made the irrelevant speech less disturbing (Item 4). The self-reported possibility of having a meeting without distracting others (item 5), working uninterruptedly for long periods (item 6), and having confidential conversations (Item 8), also all had higher scores in the masked speech condition. Lastly, participants reported that noise in their office would be less distracting if there was a water sound in the background (Item 7).

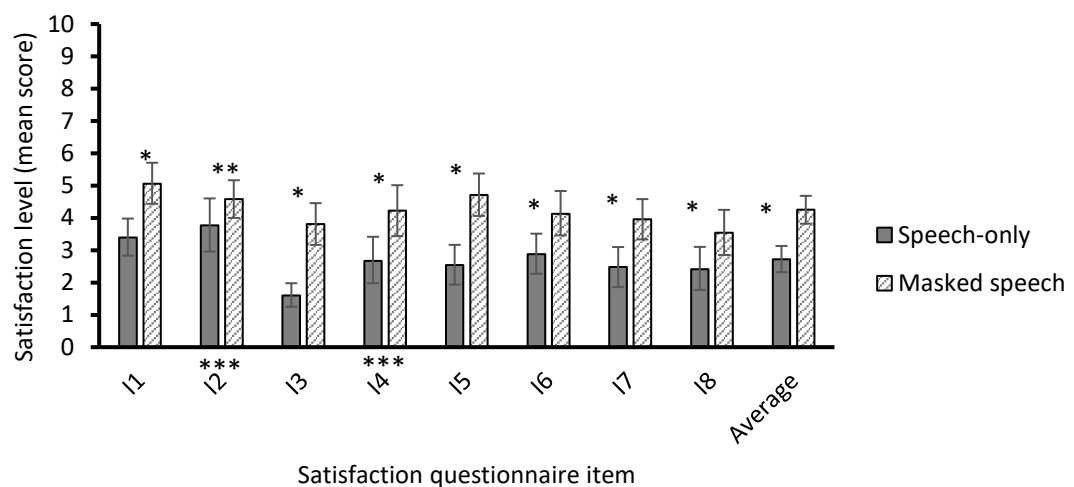


Figure 6.43 Subjective satisfaction with the overall sound environment and on the individual items' level. Error bars represent the Bias-corrected and accelerated 95% confidence intervals. Descriptions of the items are given in Table 6.2.

* $p < .001$, ** $p = .024$, ***Reverse scoring.

6.12.6 Subjective task difficulty and speech intelligibility

The self-reported difficulty of each task in both background noise conditions is shown in Figure 6.44. Ignoring the background noise conditions, the serial recall task was perceived as being the most difficult task followed by the reading comprehension task. The one-

back task was perceived as being the least difficult task followed by the information matching as the second least difficult.

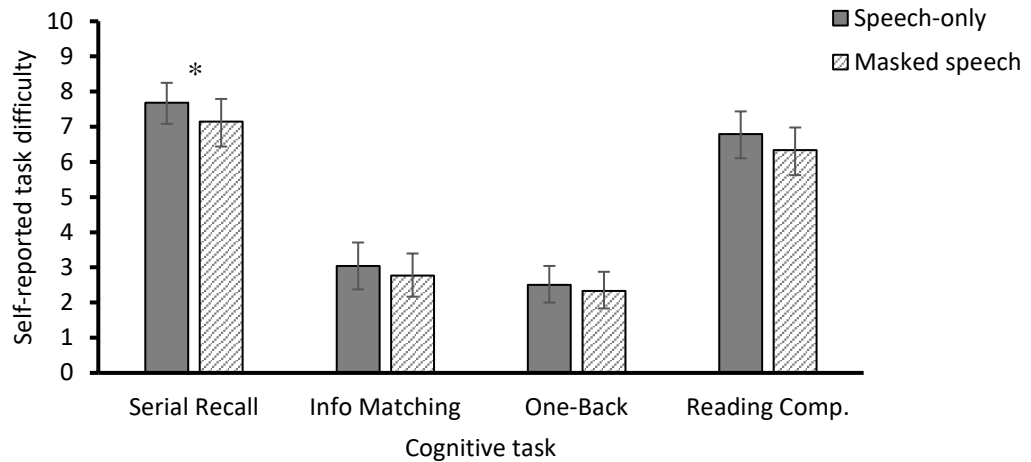


Figure 6.44 Subjective task difficulty in speech-only and masked speech conditions. Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals. * $p < .05$.

Figure 6.44 also shows that the self-reported difficulty was higher in the speech-only condition for all four tasks. This difference was statistically significant for the serial recall task, $z = -2.002$, $p = .045$, $r = -.204$; however, no statistically significant differences were detected between the self-reported task difficulty in the speech-only condition and the masked speech condition for the information matching task, $z = -1.168$, $p = .249$, $r = -.119$, the one-back task, $z = -1.001$, $p = .323$, $r = -.102$ and the reading comprehension task, $z = -1.427$, $p = .157$, $r = -.146$.

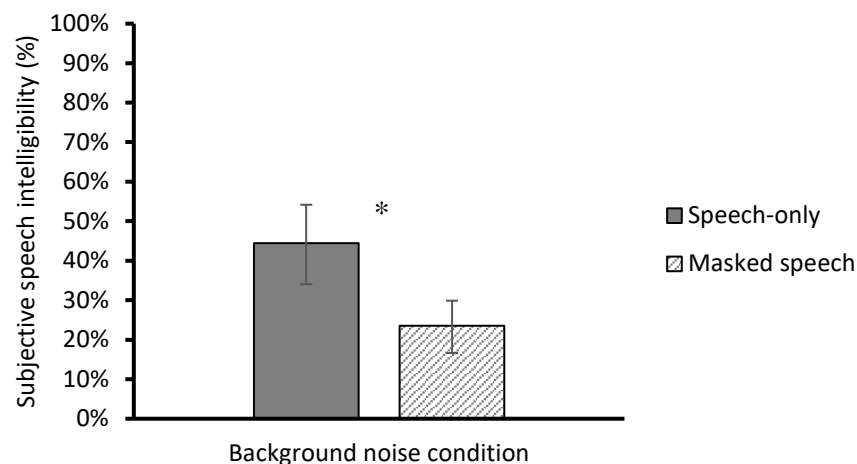


Figure 6.45 Subjective speech intelligibility in speech-only and masked speech conditions. Error bars represent the Bias-corrected and accelerated 95% bootstrap confidence intervals. * $p < .001$.

Finally, the subjective speech intelligibility in the speech-only condition was significantly higher than that of the masked speech condition, $z = -5.632$, $p < .001$, $r = -.575$. This result is shown in Figure 6.45, which supports the fact that the water sound partially masked the speech and made it less intelligible.

6.12.7 Correlation analysis

A correlation analysis was performed to further understand the relationship between the covariate, NoiseQ, and the RAs and RTs of the cognitive tasks under the two background noise conditions. In addition, the correlation analysis was used to further examine the relation between the RA and the RT of the serial recall task and the reading comprehension task. The results are presented in Table 6.4. The table shows that the covariate had a positive correlation with RT of serial recall task under masked speech condition, and RT of the reading comprehension task in both speech-only and masked speech conditions. This suggests that the more noise sensitive participants tended to take longer to complete the tasks. It is important to notice that no statistically significant correlation was found between any RA scores and the NoiseQ. Had reaction time not been measured, participants' sensitivity to noise would have been concluded to have no effect on their performance. Given the relatively simple cognitive tasks, noise sensitive and non-noise sensitive participants achieved similar performances in terms of their response accuracy. However, the more noise sensitive participants needed more time to maintain the same response accuracy than their less sensitive counterparts.

Table 6.4 Pearson's correlation coefficient (r) between selected variables measured in the task performance experiment.

Correlation between	r	BCa 95% CI		p -value
		[Lower,	Upper]	
NoiseQ and RT of serial recall task in masked speech condition	.355*	.162	.522	.017
NoiseQ and RT of reading comprehension task in speech-only condition	.314*	.128	.469	.036
NoiseQ and RT of reading comprehension task in masked speech condition	.434**	.285	.469	.003
RA and RT of serial recall task in masked speech condition	-.358*	-.616	-.081	.016
RA and RT of reading comprehension task in speech-only condition	.408**	.120	.613	.005
RA and RT of reading comprehension task in masked speech condition	.297*	.013	.524	.048

* $p < .05$, ** $p < .005$. BCa 95% CI = Bias-corrected and accelerated bootstrap 95% confidence intervals. Only statistically significant correlations are reported.

Haapakangas *et al.* (2014) found noise sensitive participants to have a lower response accuracy in a serial recall task, in comparison to their non-sensitive counterparts. In the current study, noise sensitive participants tended to take longer to complete the task. Both the current study and the study by Haapakangas *et al.* (2014) used the same NoiseQ questionnaire to capture participants' noise sensitivity. Despite adopting two different statistical methods in taking the noise sensitivity effect into account³, both studies agree that the cognitive performance of people with higher levels of noise sensitivity tend to be lower than people with lower levels of noise sensitivity.

The literature does not provide a clear-cut answer on why cognitive performance decreases as a function of noise sensitivity level. Research has observed that noise sensitive people tend to show higher levels of perceived disturbance (Haapakangas *et al.*, 2014; Pierrette *et al.*, 2014). Furthermore, Pierrette *et al.* (2014) found a positive and statistically significant correlation between noise sensitivity and annoyance level. Perhaps the more sensitive to noise a person is, the more likely they become annoyed and disturbed, which could adversely affect their cognitive performance.

The second part of the correlation analysis was to test the trade-off between response accuracy and reaction time. The general belief is that people would take a longer time to achieve a higher response accuracy and vice versa. It seems that this belief is not necessarily true, as the time-accuracy trade-off might be task dependent. For the serial recall task, there was a negative correlation between RA and RT, which indicates that the shorter the response time was, the higher the response accuracy became. Given the nature of the task, this result is not surprising. The serial recall task heavily relies on the short-term memory, as participants were required to recall digits presented on the screen shortly after they had disappeared. The longer the participants wait to recall the digits, the more chance the digits slip their minds, leading to a lower RA score.

The opposite time-accuracy trade-off was detected for the reading comprehension task, where there was a positive correlation between RA and RT, suggesting that participants with a high response accuracy tended to take a longer time to complete the task. Unlike the serial recall task, the reading comprehension task required participants to process the information presented on the screen, to be able to judge if the following statements were

³ Statistical analysis used in Haapakangas *et al.* (2014) has a discrete nature i.e., participants were either noise sensitive or non-noise sensitive, whilst in the current study, correlation analysis was adopted which has a continuous nature as it allows to input the actual noise sensitivity values of participants and look for potential trends.

true or false. Participants with a higher RA score might have taken a longer time to process the information and make a better judgement. These correlation analyses highlight the importance of considering the reaction time in any experiments measuring the cognitive performance of people. Without the reaction time, the conclusions drawn in this study would have been significantly different.

6.13 Discussion

Performance in four cognitive tasks was measured under two background noise conditions, a speech-only condition and a masked speech condition (i.e., water sound used over irrelevant speech). The tasks were a serial recall task, an information matching task, a one-back task, and a reading comprehension task. Performance in the serial recall task was significantly higher in the masked speech condition. Furthermore, this difference was significant, only when both response accuracy and reaction time were simultaneously taken into the statistical analysis (i.e., using MANOVA instead of multiple ANOVAs). For the remaining tasks, performance was not significantly different between the speech-only and the masked speech conditions. The statistical analysis used in this study allowed for taking a holistic view instead of performing analysis on individual dependent variables.

The difference in performance between male and female participants was striking and unexpected. Generally, male participants managed to maintain a steady and relatively higher level of performance in comparison to their female counterparts. The masking system did not seem to have a noticeable impact on the performance of male participants, but female participants tended to perform better in the masked speech condition. Having said that, even after the improvement in females' performance caused by the addition of the water sound, their performance was still lower than that of male participants.

It appears that the use of the water sound as a speech masking system was far more effective for females. One possible explanation is that female participants might have been too sensitive to the background speech to ignore its content and concentrate on the tasks. The addition of the masking system might have helped female participants by making the irrelevant speech less attention drawing (i.e., less intelligible) and allowing them to concentrate more on their tasks. Male participants, on the other hand, might have already developed strategies enabling them to ignore irrelevant speech, and therefore, their performance was not affected by the background noise condition. The difference between males and females within the context of cognitive performance seems to have

been overlooked in the literature. More research is needed to further advance this finding and the likely causes of these differences.

Bodin Danielsson *et al.* (2014) reported that the rate of taking short-term sick leave was higher among women, i.e., women who have been working in open-plan offices tended to take a higher number of short-term sick absences in comparison to men working in the same environment. The study suggests that women might be more vulnerable to the negative environment stimuli found in traditional open-plan offices, owing to them being more sensitive to physical stimuli of their work environment (which would include the noise environment), and hence the difference between men and women. Thus, the difference in performance between males and females in the current study can partially be justified.

Other studies have reported observing groups of high performer participants whose scores were higher than the other participants. In a study (Ebissou *et al.*, 2015) almost half of participants were insensitive to speech intelligibility, and therefore, their performance remained constant across a range of STI conditions (STI 0.25 to 0.65). The study however, did not reveal how many males and females were in each group. Therefore, it cannot be seen whether these differences in performance between the two groups could be attributed to the gender of participants. Brocolini *et al.* (2016) also identified two groups of participants, high performers and low performers, when they measured performance in a serial recall task under different background noise conditions. However, they were not able to attribute the difference in performance between the two groups to the gender nor the age of participants.

In the field of psychology observing differences in cognitive performance between males and females is common and well documented. The literature shows that men outperform women in tasks that require visuo-spatial ability (Halpern, 2000; Bosco *et al.*, 2004). Attempts have been made to explain these gender-effects either using biological factors (i.e., hormones) or genetic influences (Kimura, 1999; Mäntylä, 2013). Other studies highlighted the importance of socio-cultural factors on women's performance (Richardson, 1994; Caplan *et al.*, 1997), and observed significant effects of training and cognitive strategies in reducing these differences between men and women. Bosco *et al.* (2004) suggest that these gender effects could be due to females choosing a less effective cognitive strategy to perform visuo-spatial tasks and hence the underperformance. The study also suggests that these difference in performance between males and females could

be reduced, or even eliminated, with adequate training. On the other hand, Mäntylä (2013) suggests that menstrual changes are responsible for the difference in spatial ability between males and females. Mäntylä (2013) found females who were in the menstrual phase of the menstrual cycle to have a similar performance as males. Whilst, females who were in their luteal phase had a lower performance in comparison to males.

To some extents, the tasks examined in the current research all included visuo-spatial working memory, and hence the difference between males and females can be justified. However, it is still not clear why females benefited more from the masking sound. It could be argued that the masking sound might have helped female participants to select more efficient cognitive strategies and hence the increase in performance. However, there is no evidence to back up this claim.

Response accuracy and reaction time tended to be differently affected by the background noise condition, which once more highlights the importance of including both variables within the statistical analysis, as this allows for examining the interaction between the two. Results indicated that reaction time and response accuracy can have a negative relationship (e.g., in the serial recall task) or a positive relationship (e.g., in the reading comprehension task).

Subjectively, participants were significantly more satisfied with the masked speech condition, despite showing a significant improvement in performance (objectively) only for the serial recall task. The perceived task difficulty of all four tasks was lower in the masked speech condition, and this difference was statistically significant for the serial recall task.

There is a general agreement among scholars that speech has a detrimental effect on cognitive performance but when it comes to a specific task, there are contradictory results (refer to Table 6.1). Irrelevant speech has been reported to affect various cognitive tasks. Two explanations can be given to the discrepancies observed in reported results. Cognitive performance seems to be far more complex than the relatively simple tasks that are usually adopted to capture it, which are mainly through measuring the response accuracy or the reaction time. Using a combination of both the response accuracy and the reaction time would seem to help further understand some reasons behind the contradicting results. The second factor may have come from the over reliance on the arbitrary p -value $< .05$. Different studies have used different sample sizes and approaches to measure the performance of participants which all affect the p -value. The p -value is

affected by the sample size and thus the larger the sample size, the more chances are the differences to become statistically significant. In studies which reported no significant differences in performance between masked and unmasked speech conditions, there might have still been a genuine effect of the masking system on performance, but the sample size might have been small to make this effect statistically significant. An example would be two studies reporting results on the effect of background noise on cognitive performance, one with a $p = .051$, and the other with a $p = .049$. Practically, the difference between results reported by these two studies would be negligible, yet statistically, the former would be concluded as background noise having no effect on cognitive performance and the latter as background noise having a “significant” effect on performance. Reporting effect sizes (r) helps to overcome this problem. Unlike the p -value, the effect size is not dependent on the sample size and thus, its value is comparable between different studies. The effect size shows how large an effect is and then one can decide if its magnitude is practically meaningful, regardless of being statistically significant or not. For instance, in the analysis of the serial recall test, the effect size was $r = .310$. This is considered a *medium* effect by Cohen (1988) i.e., the effect accounts for 9% of the total variance (refer to Chapter 3). However, the effect size can be interpreted within the context of the study itself and then judged whether that effect is practically meaningful. In the case of the serial recall task, the effect size $r = .310$ means 5% increase in response accuracy and 5% reduction in response time⁴. Both figures are practically meaningful and could result in a substantial increase in productivity, and hence the profit of the company. Therefore, a medium effect size within the context of this study can be interpreted as a 5% change in performance. However, this figure should be treated with caution. It is correct only for tasks which rely on the short-term memory. Lower values should be expected for other tasks.

Combining subjective and objective measures, and taking into account the difference between males and females, it can be concluded that the water sound improved the sound environment and made people, especially females, perform some tasks better.

Limitations of this experiment include using a small room to simulate an open-plan office and playing the water sound from a loudspeaker instead of using a real water feature. In addition, to avoid making participants fatigued, the tasks were kept relatively short and therefore the longer-term effect of the water sound is not clear at this stage of the research.

⁴ This was worked out by dividing average scores of each of the RA and the RT in the masked speech condition, by their corresponding values in the speech only condition.

Visual stimuli were proven to affect the preference and perception of the water sound in the previous stage of this study. Using a real water feature might have resulted in higher performance scores in the masked speech condition, although it should be noted that all the tasks performed required looking at a screen exclusively rather than the surrounding environment.

6.14 Conclusions

This chapter looked at the effect of masking irrelevant speech on cognitive performance, and subjective satisfaction using a water sound as a speech masking sound. Four cognitive tasks were designed that represented a wide range of cognitive abilities. These tasks were a serial recall task, an information matching task, a one-back task, and a reading comprehension task. These tasks were performed by participants under two background noise conditions: a speech-only condition and a masked speech condition. The speech-only condition was created through playing a high-quality speech recording as a source of irrelevant speech. The masked speech condition was created by masking the irrelevant speech with a water sound. The water sound that was selected for this experiment was the 37-jet fountain (FTW), which was preferred in the audio-only preference test (Chapter 5). The L_{Aeq} of the water sound and irrelevant speech were set to 45 dB and 48 dB, respectively. Forty-eight participants (24 males, 24 females) took part in the experiment. Out of the selected sample, 23 participants were native English speakers (10 males, 13 females), and the remaining 25 participants (14 males, 11 females) were non-native speakers who spoke English fluently. The tests were carried out in Room G.33 of the Edwin Chadwick Building at the Edinburgh Campus of Heriot-Watt University. Cognitive performance in each task was measured through recording the response accuracy (RA) and the reaction time (RT) of participants. Hence, a multivariate approach (MANOVA) was adopted in the statistical analysis, as both RA and RT were dependent variables which represented cognitive performance. The results of the statistical analysis for each task are tabulated in Table 6.5.

Table 6.5 A summary of *F*-ratios, *p*-values, and effect sizes (*r*) obtained as a part of the statistical analysis of results in four cognitive tasks.

Tasks	Variables	<i>F</i>	<i>p</i> -value	<i>r</i>
Serial recall	RA × Bg. condition	3.100	.085	.259
	RT × Bg. condition	0.347	.559	.089
	Performance × Bg. condition	4.563	.038*	.310
	Gender	5.369	.025*	.333
	Gender × Bg. condition	1.619	.210	.190
	Gender × Performance	0.323	.573	.086
	Gender × Bg. Condition × Performance	0.055	.887	.036
	Nationality	0.332	.568	.088
	Nationality × Bg. condition	0.576	.452	.115
	Nationality × Performance	7.757	.008*	.391
	Nationality × Bg. Condition × Performance	0.020	.887	.022
	NoiseQ	1.858	.180	.204
	NoiseQ × Bg. condition	0.623	.434	.111
	NoiseQ × Performance	0.111	.954	.051
	NoiseQ × Bg. Condition × Performance	5.094	.029*	.325
	Gender × Nationality	1.534	.222	.186
	Gender × Nationality × Bg. Condition	0.200	.657	.068
	Gender × Nationality × Performance	0.581	.450	.115
Information matching	RT × Bg. condition	1.050	.311	.158
	Gender	0.148	.703	.060
	Gender × Bg. condition	0.356	.554	.093
	Nationality	6.124	.018*	.360
	Nationality × Bg. condition	1.834	.183	.207
	NoiseQ	0.380	.541	.096
	NoiseQ × Bg. condition	1.834	.276	.207
	Gender × Nationality	1.806	.186	.205
One-back	Gender × Nationality × Bg. Condition	0.075	.786	.043
	RT × Bg. condition	1.726	.196	.196
	Gender	5.346	.026*	.332
	Gender × Bg. condition	0.656	.423	.123
	Nationality	0.203	.655	.069
	Nationality × Bg. condition	0.652	.424	.122
	NoiseQ	0.073	.788	.041
	NoiseQ × Bg. condition	1.269	.266	.169
Reading comprehension	Gender × Nationality	0.001	.985	.004
	Gender × Nationality × Bg. Condition	0.321	.574	.086
	Performance × Bg. condition	0.482	.492	.108
	Gender	1.826	.184	.206
	Gender × Bg. condition	0.048	.828	.034
	Gender × Performance	0.002	.968	.007
	Gender × Bg. Condition × Performance	0.007	.935	.013
	Nationality	7.091	.011*	.384
	Nationality × Bg. condition	2.519	.120	.241
	Nationality × Performance	2.124	.153	.222
	Nationality × Bg. Condition × Performance	0.372	.545	.095
	NoiseQ	0.337	.565	.090
	NoiseQ × Bg. condition	0.063	.803	.039
	NoiseQ × Performance	3.204	.081	.269
	NoiseQ × Bg. Condition × Performance	0.419	.521	.101
	Gender × Nationality	1.027	.317	.156
	Gender × Nationality × Bg. Condition	2.307	.136	.231
	Gender × Nationality × Performance	0.181	.673	.066

* $p < .05$

Broadly speaking, cognitive performance in the serial recall task was significantly affected by the background noise condition. In comparison to the speech-only condition, participants' performance was significantly higher in the masked speech condition. For the remaining tasks, no statistically significant difference in performance was detected between the speech-only and the masked speech conditions. The analysis also revealed the importance of the combined effect of RA and RT in identifying performance. For instance, in the serial recall task, when the analysis was performed on RA alone, or RT alone (i.e., using ANOVA instead of MANOVA), no statistically significant result was observed. However, when a combined score of both RA and RT was used, the analysis revealed a statistically significant effect of the background noise condition.

In the serial recall task, the gender of participants had a significant main effect on cognitive performance. Male participants' performance score (average of RA and RT) was significantly higher than that of female participants. In addition, performance of male participants remained unaffected by the background noise conditions (similar performance scores in both conditions), while female participants' performance increased in the masked speech condition. However, this interaction was not statistically significant. Participants' nationality (i.e., whether a participant was a native English speaker or not) did not have a significant main effect on performance, but there was a significant interaction between the nationality and the performance of participants. Non-native English speakers scored a higher response accuracy, but tended to take longer to complete the task. On the other hand, native English speakers had a lower response accuracy with a shorter reaction time. There was a significant 3-way interaction between participants' noise sensitivity, background noise condition, and performance score. Under the speech-only condition, participants' RA was not affected by their sensitivity to noise, but their RT was positively correlated with their noise sensitivity (NoiseQ) score i.e., more noise sensitive participants took longer to complete the task. Under the masked speech condition, however, both RT and RA had a positive correlation with participants' NoiseQ score, suggesting that the participants who were more sensitive to noise scored a higher response accuracy, but took longer to complete the task. No further statistically significant differences were detected in the serial recall task.

In both the information matching task and the one-back task, participants' RA scores showed a ceiling effect, and hence the analysis was based on their RT scores only (i.e., using ANOVA). In the information matching task, the only statistically significant result was obtained for the nationality of participants. Native participants had a shorter reaction

time than the non-native participants. In the one-back task, the gender of participants showed a significant main effect. Similar to the serial recall task, male participants took less time to complete the task. Also, the reaction time of male participants remained unaffected by the change in the background noise condition, while, female participants showed a shorter reaction time in the masked speech condition. No further statistically significant effects were detected.

In the reading comprehension task, only the nationality of participants had a significant main effect. Native English speakers' performance was higher than that of non-native speakers. This was mainly due to non-native participants taking longer to complete the task. The RA scores of native and non-native participants were comparable. No further statistically significant results were found. The reading comprehension task included four components each measuring a different cognitive ability. The components were *text memory*, *text inferencing*, *knowledge integration* (with three levels of difficulty), and *knowledge access* (with two levels of difficulty). Participants' scores across these components were not affected by the background noise conditions.

Subjectively, overall satisfaction with the sound environment significantly increased after masking irrelevant speech with the water sound. Participants perceived the water sound to have significantly increased the pleasantness of the sound environment, and to have helped to concentrate on their tasks. The water sound made the background speech less disturbing and less attentional drawing. The possibility of working uninterrupted, having confidential conversations, and having a meeting without distracting others also significantly increased. Furthermore, subjective speech intelligibility and workload were significantly lower in the masked speech condition.

A positive correlation was found between RA and RT in the serial recall task, while a negative correlation was detected in the reading comprehension task. These results suggested that in the serial recall task, participants who took longer to complete that task, had a lower response accuracy. On the contrary, in the reading comprehension task, participants who spent more time completing the task, scored a higher response accuracy.

In the next chapter, a case study is examined which looks at the longer-term effects of exposure to a water sound in an open-plan office.

CHAPTER 7: LONGER-TERM EXPOSURE TO A WATER SOUND

7.1 Introduction

In previous experiments, the effect of masking irrelevant speech with water sounds was examined in terms of subjective satisfaction and cognitive performance. The findings of the experiments were based on participants being exposed to short periods (i.e., no more than 45 min) of speech signals masked by water sounds. Furthermore, findings obtained in Chapter 5 supported the use of a real water feature over playing its sound through speakers. To further extend these findings, an additional experiment was carried out, to examine the impact of longer-term exposure to a water sound in an open-plan office. This chapter presents methodologies, results and findings of this experiment. The experiment included installing a water feature in an open-plan office for a period of three weeks. A satisfaction questionnaire was distributed before and after the installation of the water feature. The responses given before and after the installation of the water feature were compared to examine how people's perception changed after the inclusion of the water feature in the space. The chapter starts by providing descriptions of the water feature used in the study, and the space in which the water feature was placed. Descriptions of the participants and the questionnaires are then provided, followed by results that are presented, analysed and critically discussed. The main findings of the experiment are summarised at the end of the chapter followed by the limitations of this experiment.

7.2 Design of the water feature

A water feature (Daintree Planter Cascade Water Feature, dimensions, $0.48\text{H} \times 0.45\text{W} \times 0.49\text{D}$ m) was purchased from an online retailer (primrose.co.uk), and modified to meet the criteria set out in the previous stages of this research. A cascade-like water feature was highly preferred in the audio-visual preference test, and therefore, a 3-step cascade was purchased and modified to have a pleasant sound and look. Modifications included replacing the water pump to have a higher flow rate (1300 LPH), and adding pebbles as impact material instead of the thin plate from which the water feature was built. The water pump provided with the water feature had a relatively low flow rate (700 LPH) which was incapable of generating a high enough sound pressure level. Although

the current study mainly concentrates on the audio aspects of the water feature, in the audio-visual preference and perception tests (Chapter 5), it was found that the visual stimulus has a positive effect on the people's perception; therefore, the water feature in the current experiment was painted in white (came in black originally) and cactus plants were added to the planters provided with the water feature, in order to improve its aesthetic value. The modified water feature is shown in Figure 7.1.



(b) Left view



(a) Right view

Figure 7.1 The 3-step cascade used in Experiment 4 of the current study.

After modifications, the sound quality of the water feature was subjectively evaluated and the sound pressure level of the water feature and its directivity were measured in the highly insulated anechoic chamber of Heriot-Watt University. The energy averaged sound pressure level, 1 metre away from the centre of the water feature was measured to be 45.5 dBA. This average number is the result of measuring the SPL at eight positions around the water feature and 1 position above it, each 1 m away from the centre of the water feature. The directivity of the water feature, and the positions at which the sound pressure levels were measured are shown in Figure 7.2.

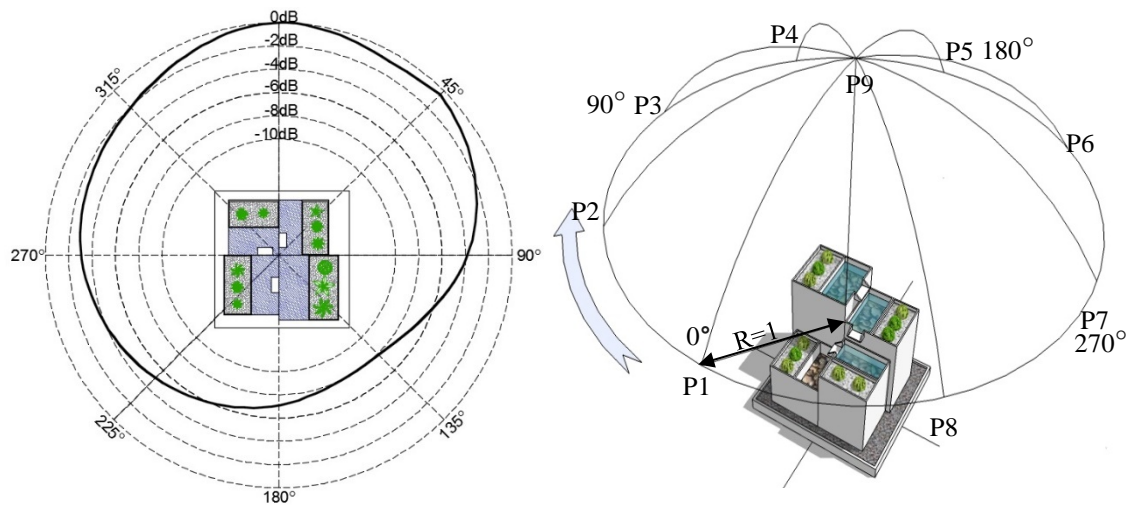


Figure 7.2 Directivity (left) and positions at which the sound pressure levels of the water feature was measured (right) in the anechoic chamber. The outer circle of the directivity diagram represents 48.5 dBA measured at position 1 (P1), 1 m from the centre of the water feature.

The open-plan office

A small-sized open-plan office with a floor area of 56.3 m² (dimensions, 7.60W × 8.75L × 2.90H m) was selected for the water feature to be installed in. Figure 7.3 shows the plan of the office with its dimensions, and Figure 7.4 shows a photograph of the space. The open-plan office was in the William Arrol Building of the Edinburgh campus of Heriot-Watt University. The office accommodated 12 workstations clustered into 3 groups of working areas. The finishing material of the walls was plaster and the ceiling was made of absorbent ceiling tiles. The water feature was placed at the middle of the shorter side of the space on a 0.7 m high table as shown in Figure 7.4. This position was carefully chosen to minimise the space taken up by the water feature while making it both visible and audible from most of the workstations. The office was located in a quiet area far from road traffic and construction noises. The equivalent sound pressure level, L_{Aeq} , of the background noise in the absence of employees at each workstation was measured over a period of 15 seconds, with and without the water feature in operation. The space averaged $L_{Aeq,15s}$ was 33.5 dB (empty room but equipment switched on) without the water feature, and rose to 39.3 dB when the water feature was switched on. The octave band spectra of the background noise with and without the water sound are shown in Figure 7.5.

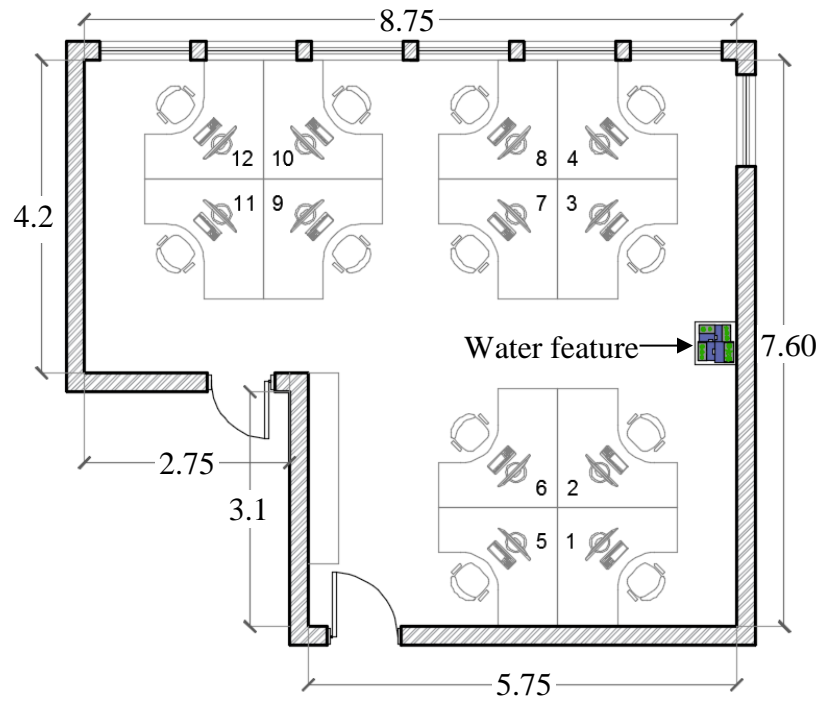


Figure 7.3 Plan of the open-plan office where the water feature was placed in. Dimensions are given in metres.



Figure 7.4 A photograph of the open-plan office where the water feature was placed in.

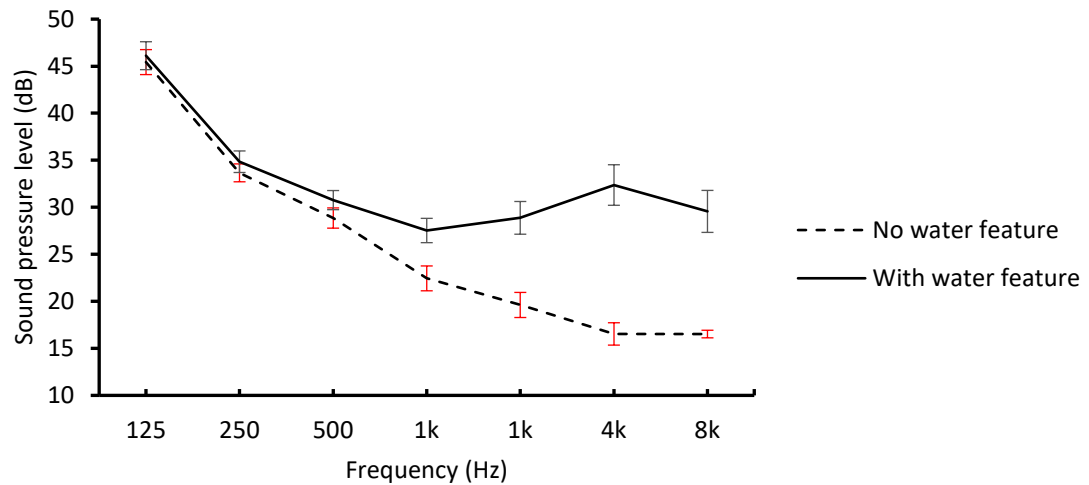


Figure 7.5 Octave-band spectra of the background noise with and without the water sound, space averaged over 12 workstations. Error bars represent the 95% confidence intervals.

7.3 Questionnaire

A modified version of the GABO questionnaire (Pierrette *et al.*, 2014) was used to assess the initial noise environment in the open-plan office and to measure likely effects that the water feature could have on people’s perception of their work environment. The questionnaire was divided into two parts, Part 1 and Part 2. Part 1 was distributed before installing the water feature, and Part 2 was distributed after the water feature had been in the space for a period of 3 weeks. The two parts were mostly identical apart from a few extra questions concerning the water feature added in Part 2. Both Part 1 and Part 2 of the questionnaire are provided in Appendix J and Appendix K, respectively.

The questionnaire was based on bibliographic research and semi-directive interviews carried out by Pierrette *et al.* (2014), and is structured around four main sections. Section 1, “General information about you and your workstation”, gathered participants’ background information such as age, gender and length of time working in the office. This section also included a statement concerning participants having any known hearing difficulties. Section 2 of the questionnaire, “Assessing the physical environment of your work area” assessed the employees’ satisfaction with their physical working environment. This section consisted of 14 items, half of which (i.e., 7 items) measured satisfaction relating to Control/Privacy aspects and the other half was about Comfort/Functionality aspects of the workspace. This section resulted in three satisfaction scores; a global satisfaction level based on the average score of all 14 items, a Control/Privacy satisfaction and a Comfort/Functionalities satisfaction, each based on the average score of 7 items.

Section 3, “Assessing the noise environment of your work area”, assessed employees’ noise environment. Participants rated the general perceived noise level and then stated the level of annoyance caused by noise in the space. Participants were also asked about the perceived frequency of occurrence of two noise sources: intelligible speech and unintelligible speech. The level of annoyance caused by the noise sources, the impact of the noises on work performance, and whether some specific tasks were sensitive to the noises were also measured. Section 3 also included a question which required participants to rank 8 noise sources from most annoying to least annoying. The noise sources covered in this question were:

- ventilation/air-conditioning noise;
- printers/photocopiers;
- telephones ringing;
- conversations in which you can hear but cannot understand;
- conversations which you can understand what is said;
- people walking up and down the office;
- noise of people working; and
- noise linked to one person in particular.

Section 4, “Your perception of the sound environment”, was dedicated to measuring people’s perception of the sound environment in the working space. This section did not exist in the original questionnaire and was added in the current version to allow for measuring the likely changes in people’s perceptions of their sound environment associated with installing a water feature. The section included 6 questions and measured different aspects of the sound environment such as pleasantness, possibility of concentrating on tasks, possibility of having meetings without distracting others, working uninterrupted, and possibility of having private conversations.

Ten more questions specific to the water feature were added to this sections in Part 2 of the questionnaire, 6 of which were given as statements, in addition to three Yes/No questions specific to the water feature plus an open-ended question. The statements were “The water feature is pleasant”, “The water feature has improved the sound environment”, “The water feature helps me to carry out private conversations”, “The water feature stresses me”, “The water sound is distracting”, and “The water feature is visually/aesthetically pleasing”. The three Yes/No questions were “Does the water feature help you to carry out a particular task in a better way?”, “Does the water feature make

you go to the toilet more frequently?”, “Would you like the water feature to remain in your office permanently?”. The open-ended question asked participants if there was anything else they would like to address that had not been covered by the questionnaire.

Section 5, “Your relationship with noise in general” assessed how people reacted to noise, i.e., their sensitivity to noise. This section was a shorter version of the noise sensitivity questionnaire (NoiseQ) developed by Schütte, Sandrock, *et al.* (2007), which consists of 12 items divided into 3 subscales, namely, *sleep*, *habitation* and *work*, with 4 items in each subscale. Participants stated their level of agreement with each item on a 4-point numerical scale with 1 representing “strongly disagree”, and 4 representing “strongly agree”. The answer to each question was then quantified from 0 to 3 and used to calculate the average noise sensitivity score. A score of less than 1.11 is considered as not being sensitive to noise, while a score greater than 1.63 is considered as being sensitive to noise (Schütte, Marks, *et al.*, 2007). The original GABO questionnaire included an extra section concerning perceived health of participants, but it was omitted in the current version for being of little use to the current study and not to further extend the questionnaire.

In the original questionnaire, a 5-point Likert scale was adopted, apart from the noise sensitivity section for which a 4-point scale was used. However, an 11-point numerical scale was used in the current study where 0 stood for “very dissatisfied/strongly disagree”, and 10 stood for “very satisfied/strongly agree”. The main aim of the study was to examine the likely effects that installing a water feature has on people’s satisfaction and perception of their work environment, which tend to be small, and small differences are more easily detectable on an 11-point scale than, for example, a 5-point Likert scale.

7.4 Participants

Fourteen participants (2 males, 12 females) filled out the questionnaires. These were staff members of Heriot-Watt University, aged between 24 and 61 *yr* ($M = 39.86$ *yr*, $SD = 11.64$ *yr*). The average time participants had spent in the open-plan office was $M = 1.39$ *yr* ($SD = 1.15$ *yr*), with an average attendance of $M = 2.86$ days per week ($SD = 1.55$ days per week), due to staff rotation.

7.5 Statistical analysis

The data was analysed with IBM SPSS Statistics for Windows, Version 22.0. Given the small sample size ($N=14$) and the violation of the assumption of normality of most scores

(checked using Shapiro-Wilk test and Normality Q-Q plot), it was decided to adopt non-parametric tests for the statistical analysis. The Wilcoxon signed-rank test was used for comparing scores between the two parts of the questionnaire. Spearman's correlation coefficient, r_s , was used for the correlational analysis between variables. The internal consistency of scores within each section of the questionnaire was assessed using Cronbach's alpha, α . The Bias-corrected and accelerated bootstrap method, BCa, was used to derive robust 95% confidence intervals (CI), which are reported in square brackets throughout this chapter.

Where appropriate, the effect size, r , is given, which is a standardised measure of the size of effect observed. The effect size is not readily available in SPSS, however, the z -scores provided as a part of the Wilcoxon signed-rank test can be converted to r , using the following equation (Field, 2013):

$$r = \frac{z}{\sqrt{N}} \quad (7.5)$$

where z is z -score from the Wilcoxon signed-rank test, and N is number of observations, which in this study is equal to twice the number of participants, i.e., 28 (14 x 2). No comparison was made between scores of male and female participants due to having only 2 males in the sample.

7.6 Procedure

Part 1 of the questionnaire measured the initial satisfaction level of workers within their work environment. After all participants had filled out Part 1, the water feature was installed in the space. The water feature remained in the office for 3 weeks (5 days/week), before Part 2 of the questionnaire was distributed. Participants were asked to keep Part 1 until after Part 2 of the questionnaire was distributed. Then, both parts were collected together. The responses obtained from both parts of the questionnaire were analysed and compared to identify any change in people's satisfaction level and perception of the work environment.

7.7 Results

Participants' responses to all items included in Part 1 and Part 2 of the questionnaire are tabulated in Appendix L.

7.7.1 Participants' sensitivity to noise

The score obtained from the noise sensitivity questionnaire revealed that, on average, participants were moderately sensitive to noise ($M = 1.58$, $SD = 0.70$). A slightly higher score was obtained when the sensitivity to noise was calculated based on the subscale *work* with an average score of $M = 1.66$ ($SD = 0.64$). The reliability analysis revealed a very high internal consistency, with a Cronbach's $\alpha = .928$, which is slightly higher than the Cronbach's $\alpha = .84$ reported by Pierrette *et al.* (2014). When only the 4 items of subscale *work* were included in the reliability analysis, the internal consistency dropped to Cronbach's $\alpha = .860$, which is understandable, as fewer items were included in the analysis and Cronbach's α is known to be affected by the number of items (Field, 2013).

7.7.2 Satisfaction with the physical work environment

The results from Part 1 of the questionnaire revealed that participants were satisfied with their physical work environment, $M = 7.06$ [6.55, 7.64]. When the analysis was made separately for the two underlying subscales, the results indicated that participants were more satisfied with the comfort/functionality aspects of their work environment, $M = 7.58$ [7.09, 8.09], compared to control/privacy aspects, $M = 6.53$ [5.88, 7.24]. The difference in satisfaction between the two subscales was statistically significant, $z = 2.984$, $p = .001$, $r = .564$.

After the water feature was added to the space, the global satisfaction level within the physical work environment increased, $M = 7.28$ [6.71, 7.92], however, this increase was not statistically significant ($z = 1.615$, $p = .115$, $r = .305$). Having said that, the water feature significantly increased the satisfaction levels within the subscale comfort/functionality, $M = 7.84$ [7.35, 8.36] ($z = 2.530$, $p = .012$, $r = .478$). The satisfaction level within the subscale control/privacy also increased, $M = 6.71$ [5.91, 7.56], but the increase was not statistically significant ($z = 1.104$, $p = .295$, $r = .209$). Figure 7.6 and Figure 7.7 show people's responses to items included in subscales comfort/functionality, and control/privacy, respectively.

The increase in satisfaction level can be attributed to 3 items within the physical work environment, namely *Item 1* "Noise environment" and *Item 2* "The cleanliness of your work area" within the subscale comfort/functionality, and *Item 3* "Possibility of concentrating in your workplace" within the subscale control/privacy. When the scores of Items 1, 2 and 3 were compared before and after installing the water feature, it was revealed that the water feature significantly increased the satisfaction level for *Item 1* ($z = 2.803$, $p = .004$, $r = .530$), and *Item 2* ($z = 2.041$, $p = .041$, $r = .386$), but the increase was not statistically significant for *Item 3* ($z = 1.876$, $p = .074$, $r = .354$).

These results suggest that the water feature significantly improved the noise environment and made participants perceive their environment to be cleaner. There was no cleaning process before and after installing the water feature, therefore improvement in the cleanliness of the working space was unexpected. This improvement could be attributed to the water feature increasing the aesthetic value of the space, which could have been interpreted by participants as a cleaner environment. The improvement might also be attributed to the fact that water is normally associated with cleansing and washing.

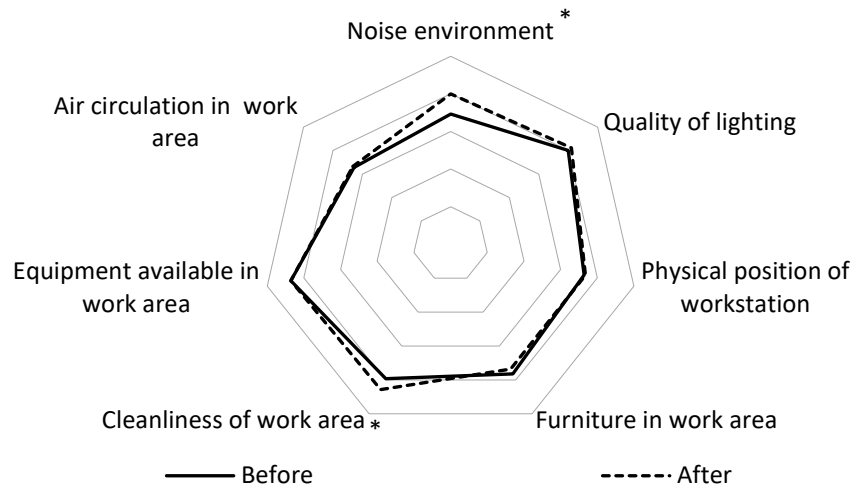


Figure 7.6 Satisfaction level within Comfort/Functionality aspects of the physical work environment, before and after installing the water feature. Centre of polygons is Zero, outer polygon is 10. * Difference is significant at the .05 level.

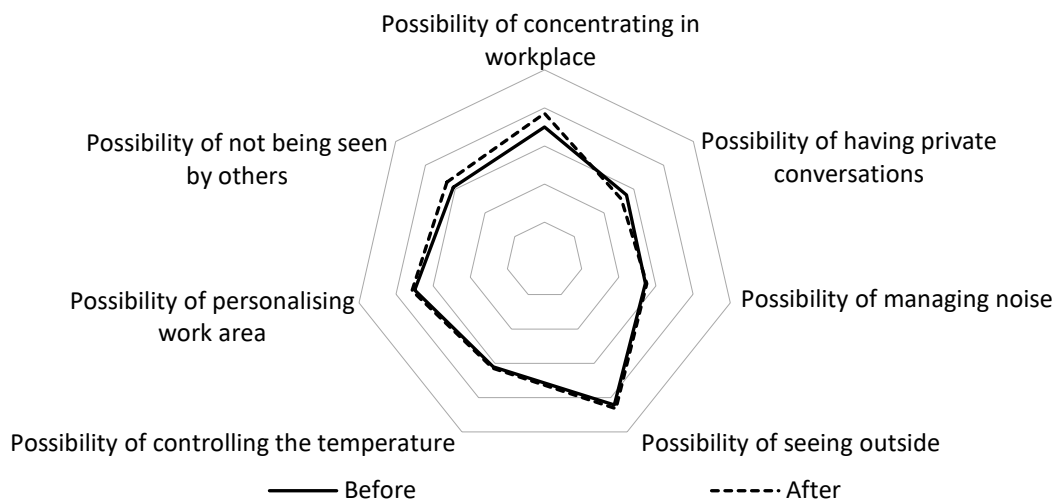


Figure 7.7 Satisfaction level within Control/Privacy aspects of the physical work environment, before and after installing the water feature. Centre of polygons is Zero, outer polygon is 10.

Generally, the water feature improved the Comfort/Functionality aspects of the space, however, people were already satisfied with the physical work environment before the

inclusion of the water feature and it is not clear if the water feature would have resulted in a similar change had the initial satisfaction level been low.

The physical work environment section of both parts of the questionnaire had high reliabilities, with Cronbach's $\alpha = .789$ and $\alpha = .843$, before and after installing the water feature, respectively. However, lower reliabilities within the subscales Comfort/Functionality (Cronbach's $\alpha = .565$), and Control/Privacy (Cronbach's $\alpha = .704$) were detected, for Part 1 of the questionnaire. In Part 2 of the questionnaire, reliabilities within the subscales increased to Cronbach's $\alpha = .603$ for Comfort/Functionality, and Cronbach's $\alpha = .826$ for Control/Privacy.

7.7.3 Satisfaction with the noise environment

The satisfaction level of participants for six items included in the noise environment section is shown in Figure 7.8. Employees perceived the noise level, $M = 4.43$, 95% [3.14, 5.97], and the annoyance level, $M = 4.14$ [3.14, 5.21], to be moderate. Furthermore, the installation of the water feature did not seem to have any effect on perceived noise level, $M = 4.29$ [2.93, 5.64], and annoyance, $M = 4.36$ [3.21, 5.43]. Statistically, no significant differences in perceived noise level, $z = 0.426$, $p = .672$, $r = .081$, and annoyance, $z = -0.412$, $p = .680$, $r = -.078$, were detected before and after installing the water feature.

Participants assessed the frequency of occurrence of two noise sources, intelligible speech and unintelligible speech. Intelligible speech, $M = 7.50$ [6.64, 8.36], was perceived as being twice as frequent as unintelligible speech, $M = 3.79$ [2.14, 5.36]. After the water feature was added, the perceived frequency of intelligible speech dropped to $M = 6.21$ [5.36, 7.00], and the change was statistically significant, ($z = -2.326$, $p = .027$, $r = -.440$). On the contrary, the water feature resulted in an increase in the perceived frequency of unintelligible speech, $M = 4.57$ [3.14, 5.93], yet, the increase was not statistically significant ($z = 0.784$, $p = .523$, $r = .148$). This is understandable, as the water sound must have masked a portion of the intelligible speech and made it unintelligible, hence the increase in the perceived frequency of unintelligible speech. Looking at the r values, it appears that the water sound had a much larger effect on reducing the frequency of occurrence of intelligible speech in comparisons to its effect on increasing frequency of occurrence of unintelligible speech. This implies that in addition to masking a portion of the intelligible speech and making it unintelligible, another portion of the intelligible speech might have become inaudible, hence the inequality in effect sizes.

In terms of annoyance, neither intelligible speech, $M = 3.93$ [2.57, 5.29], nor unintelligible speech, $M = 2.43$ [1.21, 3.86], was perceived as being excessively annoying. Furthermore, the water feature did not have a significant impact on reducing the perceived annoyance level caused by intelligible speech, $M = 3.86$ [2.64, 5.00] ($z = -0.276$, $p = .783$, $r = -.052$), and resulted in an increase in the annoyance level associated with unintelligible speech, $M = 3.50$ [2.07, 5.07], yet the increase was not statistically significant ($z = 1.689$, $p = .109$, $r = .376$).

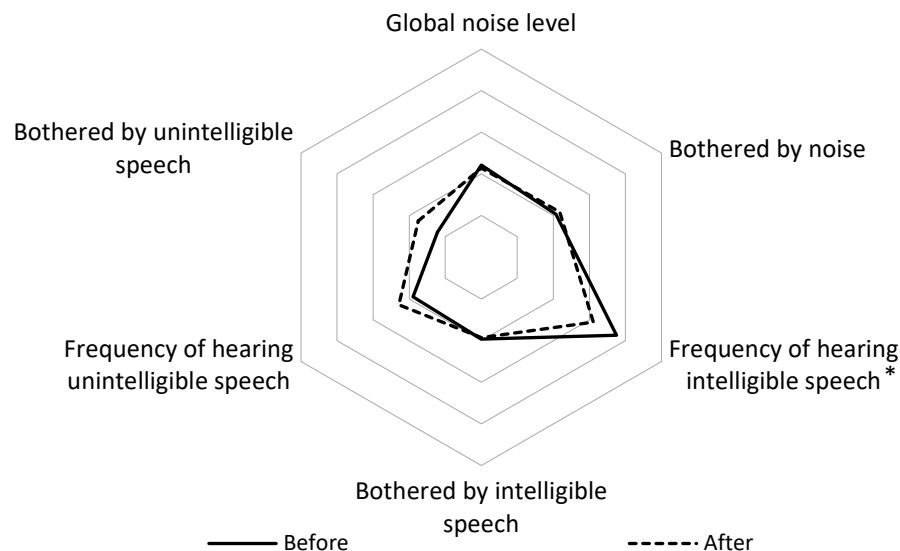


Figure 7.8 Satisfaction with noise environment before and after installing the water feature. Centre of polygons is Zero, outer polygon is 10. * Difference is significant at the .05 level.

When participants were asked if there was any task for which intelligible speech was particularly distracting, 7 out of 14 answered “yes”, and the tasks mentioned were writing (4 times), reading, data entry, record updating, and telephone conversations (2 times each). For unintelligible speech, only 4 participants thought that unintelligible speech was particularly distracting for certain tasks, and the tasks were similar to those highlighted above. The presence of the water feature did not seem to have any effect on the type of tasks affected by intelligible and unintelligible speech (see Appendix L).

Regarding intelligible speech, 8 participants claimed perceived annoyance to be the same regardless of whether they heard both sides of a conversation or just one side (i.e., telephone conversation). Three participants were more annoyed by two-sided conversations and the remaining three were more annoyed by one-sided conversations. These results partially support both the *Irrelevant Speech Effect* (Salamé and Baddeley,

1989) and the *Changing-state Hypothesis* (Jones *et al.*, 1992), as both one-sided and two-sided conversations were perceived to be equally disturbing (see Section 2.10.2). Concerning the preferred working space, an open-plan office and a shared office (shared with 1 to 3 people) were equally preferred by 6 participants each. A private office was only preferred by 2 people. The office examined in this experiment was for administrative staff who provide support for postgraduate research students. The job requires a degree of interactions among the staff members, which is possibly why the shared office and the open-plan office were preferred over the private office.

Participants were also asked to rank eight noise sources from most annoying to least annoying. Figure 7.9 shows the ranking for all eight noise sources before and after installing the water feature. Before the installation of the water feature, the most annoying source of noise was found to be “telephone ringing”, followed by intelligible speech and unintelligible speech, respectively. The least annoying noise source was “ventilation/air-conditioning noise”, followed by “noise linked to one person in particular”, and “printers/photocopier”. Work-related noise (i.e., key board, opening and closing drawers, etc.) and noise made by “people walking up and down the office” were moderately ranked between most and least annoying noise sources, with the work-related noise being perceived as slightly less annoying than the noise made by “people walking up and down the office”.

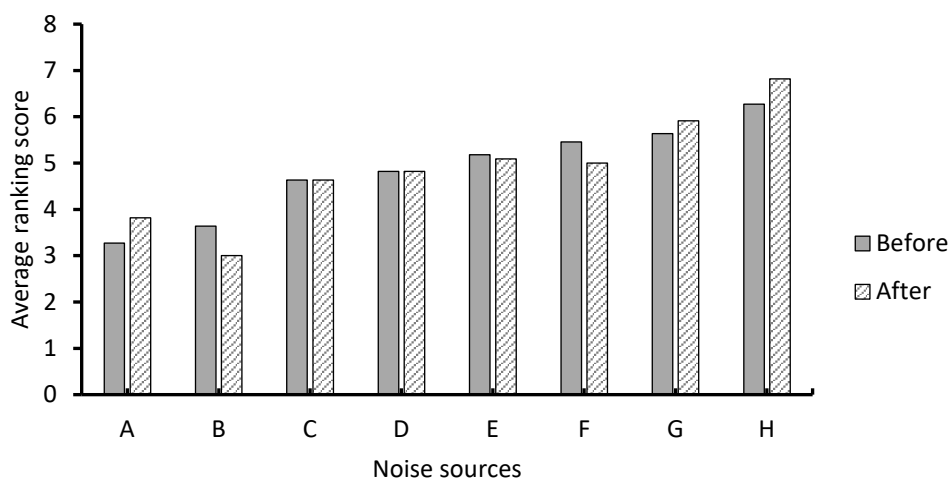


Figure 7.9 Average ranking scores for 8 noise sources according to their annoyance level. A = telephones ringing, B= intelligible speech, C = unintelligible speech, D = people walking up and down the office, E = work-related noise (keyboard, opening and closing drawers, etc.), F = Printers/Photocopier, G = noise linked to one person in particular, and H = Ventilation/Air-conditioning noise. Higher scores mean less annoying noise source.

The ranking did not change dramatically after installing the water feature. The most annoying noise source became “intelligible speech” followed by “telephone ringing”, and “unintelligible speech”, respectively. Noise made by “people walking up and down the office” was perceived as the fourth most annoying noise source (i.e., same as before), followed by “printers/photocopier noise”, and work-related noises. “Ventilation/air-conditioning noise” remained at the bottom of the ranking, followed by “noise linked to one person in particular”.

7.7.4 Perception of sound environment

In the previous section “Satisfaction with the noise environment”, annoyance levels caused by different noise sources were addressed, while this section is more focused on the pleasantness aspects of the sound environment. This part of the questionnaire included six questions, two of which examined people’s perception of the sound environment, while the remaining four were related to the possibility of carrying out certain office-related activities within the sound environment. As shown in Figure 7.10, the water feature resulted in an increase in score for all 6 questions.

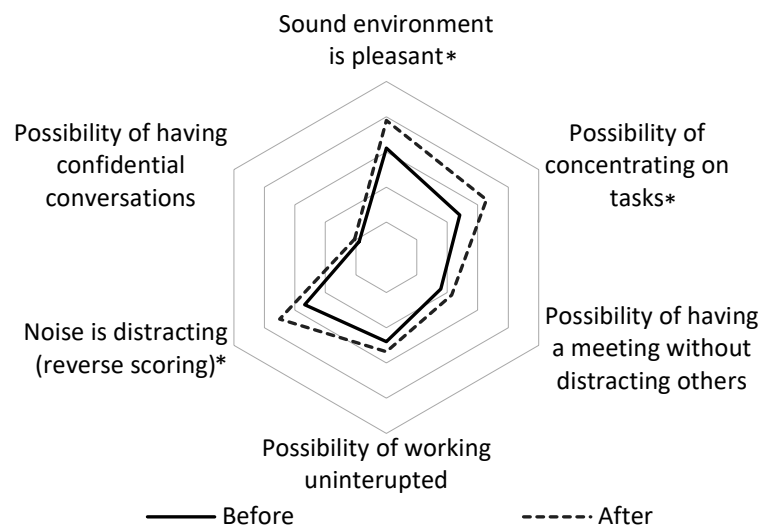


Figure 7.10 Satisfaction level with the sound environment, before and after installing the water feature. Centre of polygons is Zero, outer polygon is 10. * Difference is significant at the .05 level.

Participants perceived the sound environment of the open-plan office to be moderately pleasant, $M = 6.21$ [5.21, 7.14]. The inclusion of the water feature in the work space significantly increased the pleasantness level, $M = 7.79$ [6.86, 8.71] ($z = 3.244$, $p < .001$, $r = .613$). The sound environment was perceived as not being particularly effective in helping people to concentrate on their tasks, $M = 4.82$ [4.07, 5.61], but this score

significantly increased after the installation of the water feature, $M = 6.57$ [5.57, 7.64] ($z = 2.807, p = .003, r = .530$). This suggests that the water feature was effective in helping people to carry out certain tasks. Participant were also asked if they perceived the sound environment of the work space to be distracting, and the responses suggest that they did not perceive it as being distracting before the water feature was installed, $M = 4.64$ [3.71, 5.64]; the water feature further improved the sound environment and made it significantly less distracting, $M = 3.00$ [1.93, 4.07], ($z = -2.505, p = 0.014, r = -.474$).

Despite slight increases, no significant differences were detected in the possibility of having a meeting without distracting others ($z = 1.111, p = .344, r = .210$), the possibility of working uninterrupted for long periods ($z = 0.905, p = .563, r = .171$), and the possibility of having confidential conversations ($z = 0.496, p = .664, r = .094$). The latter was to be expected, as confidential conversations require a greater level of privacy that the water sound was incapable of providing. A private office would probably be more suitable for this type of conversations.

The questions included within this section of Part 1 of the questionnaire had a very low internal consistency, Cronbach's $\alpha = .378$. However, for the same set of questions, but after installing the water feature, (i.e., Part 2 of the questionnaire) the internal consistency significantly increased to score a Cronbach's $\alpha = .841$. The lower Cronbach's α was taken into account in the analysis, as it suggested that the questions might not have measured the same aspect of the sound environment, and therefore, the analysis was based on the individual score of each item instead of an average score of all questions included within this section of the questionnaire.

Table 7.1 Items added to Part 2 of the questionnaire concerning the water feature, and participants' responses to those items.

Questions/Statements	Mean score
1 The water sound is pleasant.	9.14 [8.64, 9.57]
2 The water feature has improved the sound environment.	8.57 [7.93, 9.14]
3 The water feature helps me to carry out private conversations.	4.43 [3.07, 5.86]
4 The water feature stresses me (<i>reverse scoring</i>).	0.14 [0.07, 0.36]
5 The water sound is distracting (<i>reverse scoring</i>).	0.14 [0.07, 0.36]
6 The water feature is visually/aesthetically pleasing.	9.57 [9.29, 9.86]
7 Does the water feature help you to carry out a particular task in a better way?	No = 11, Yes = 3
8 Does the water feature make you go to the toilet more frequently?	No = 12, Yes = 2
9 Would you like the water feature to remain in your office permanently?	No = 1, Yes = 13
10 Open-ended question	N/A

Note: Bias corrected and accelerated bootstrap 95% CIs are reported in square brackets.

Part 2 of the questionnaire included 10 extra items concerning the water feature, as mentioned before. These questions were added to examine any negative impact that the water feature and its sound might have had which was not addressed by previous items. Six questions were given as statements, three more questions required participants yes/no answers and if “yes” was selected, there was an option for clarification. These questions and the responses given by participants are tabulated in Table 7.1.

The responses to these questions revealed that the water sound was very positively perceived as being pleasant, $M = 9.14$ [8.64, 9.57], as improving the sound environment, $M = 8.57$ [7.93, 9.14], and as being visually/aesthetically pleasing, $M = 9.57$ [9.29, 9.86]. Furthermore, the water feature did not cause people to feel stressed, $M = 0.14$ [0.07, 0.36], nor did its sound distract people, $M = 0.14$ [0.07, 0.36]. Nevertheless, the water feature did not seem to have helped people to carry out private conversations, $M = 4.43$ [3.07, 5.86], which confirms the response obtained within the previous section of the questionnaire when a similar question was asked but in a less direct way.

When participants were asked if the water feature had helped them to concentrate on a particular task, three participants answered “yes” and the remaining 11 participants answered “no”. This result somehow contradicts a result reported in the previous section, when the water feature was found to have significantly increased the possibility of concentrating in tasks (see Figure 7.10). It appears that the scale used in the questionnaire, had an influence on people’s response. When an 11-point numerical scale was used, the result showed that the addition of the water feature significantly increased the possibility of concentrating on tasks, while the binomial scale (i.e., yes or no) did not reveal a significant improvement. In the case of the 11-point numerical scale, people were given the opportunity to state the magnitude of improvement, while in the binomial scale, the magnitude cannot be stated, which could have made people reluctant to select “yes”, simply because of them being unable to express how beneficial the water feature was. Only two participants thought that the water feature increased the frequency of going to the toilet with an average increase of two times per a day. Finally, 13 out of 14 participants preferred the water feature to remain in the space on a permanent basis.

The last question was an open-ended question asking people to address anything else concerning the water feature that was not covered before. The comments were all positive and some of them are quoted; “Very pleasant addition to the office, it would be nice if it could stay. However, I think more would be required to improve the environment”, “I don't think it masks other sounds (not that I am aware of), but it is a pleasant soothing

sound. I think it improves the office ambience”, “I find it difficult to work in silence, so whilst I do not feel it has improved our ability to have private conversations, it has definitely benefitted me personally by breaking the silence and improving concentration. I would like the feature to remain as the water sound is relaxing and in no way distracting”, “The feature is lovely to look at, the sound is pleasant, but it does not really counteract the noise we all make in this office! When it is turned off at the end of the day, it is almost too quiet!”. The responses to the above questions suggest that the water feature was highly appraised by participants with a very little adverse effect on the number of times that people needed to go to the toilet.

7.7.5 Correlational analysis

Four independent variables were measured and used to find possible correlations with the dependent variables (i.e., scores given to each item or subscales of the questionnaire). The independent variables were age of participant, overall noise sensitivity (NoiseQ overall), noise sensitivity for the subscale “Work” (NoiseQ Work) and the equivalent background noise level at each workstation, $L_{Aeq,15s}$, with the water feature in operation. These independent variables were compared to participants’ responses to different items in Part 2 of the questionnaire. Table 7.2 shows Spearman’s correlation, r_s , between the independent and some of the dependent variables. Due to the relatively small sample size ($N = 14$), finding a significant correlation was unlikely, but the size of the correlation provides more information than the p -value (Field, 2013). A cut-off point of $r_s = .50$ was used, as it represents a sensible relationship between two variables, and therefore, correlations smaller than the cut-off point are not reported. The table also presents the p -values and the Bias-corrected and accelerated 95% bootstrap confidence intervals of r_s . If the confidence intervals do not include zero (i.e., both lower and upper bound remain positive or negative), then the correlation is more likely to be genuine and reliable (Field, 2013).

The age of participants was found to positively correlate with the level of satisfaction with the *physical position of work station* ($r_s = .598, p = .024$), the *possibility of not being seen by others* ($r_s = .526, p = .053$), and the *possibility of meeting without distracting others* ($r_s = .581, p = .029$). The age of participant was negatively correlated with the *perceived frequency of occurrence of unintelligible speech* ($r_s = -.509, p = .063$), suggesting that the perceived frequency of occurrence of unintelligible speech was lower for older participants.

Table 7.2 Spearman's correlation (r_s) between participants' responses and their age, noise sensitivity, and background noise level at workstations. r_s values are shown in bold.

Questions/Subscales	Age	NoiseQ Overall	NoiseQ Work	B. Noise level
Satisfaction with physical position of work station. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit	.598* .024 .155 .898			
Possibility to personalise work area. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit				.583* .036 .053 .904
Possibility of not being seen by others. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit	.526 .053 -.044 .898			
Satisfaction with Comfort/Functionality Subscale. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit				.570* .042 .110 .822
Perceived level of global noise. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit			.614* .019 .115 .905	
Annoyance caused by perceived global noise. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit			.654* .011 .250 .879	
Annoyance caused by intelligible speech. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit		.662** .010 .371 .809	.760** .002 .474 .913	-.605* .028 -.951 .039
Frequency of occurrence of unintelligible speech. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit	-.509 .063 -.883 .114			
Possibility of meeting without distracting others. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit	.581* .029 .051 .910			
Possibility of working uninterrupted. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit				.652* .016 .151 .878
Distraction caused by noise in the office. <i>p</i> (2-tailed) <i>BCa</i> %95 CI, Lower limit <i>BCa</i> %95 CI, Upper limit			.556* .039 -.090 .869	-.744** .004 -.914 -.317

*Correlation is significant at the 0.05 level.

**Correlation is significant at the 0.01 level.

Two other possible explanations for these correlations are Presbycusis and the U-shaped happiness curve. Presbycusis is a common hearing loss caused by the natural aging of the hearing organs. In the current study, older participants might have showed a high satisfaction level simply because of having a less sensitive hearing which could have made them less sensitive to irrelevant speech. The U-shaped happiness curve, on the other hand, is a trend which shows relationship between happiness and age. Blanchflower and Oswald (2017) reviewed seven studies covering 51 countries and nearly 1.3 million sampled people to establish a correlation between life satisfaction, or happiness and age. They established a U-shaped trend between age and happiness, with middle age people showing a lower level of satisfaction in comparison to younger and older people. Taking this U-shaped happiness curve into account, for the same environment, elderly people tend to be more satisfied than middle aged people.

The overall noise sensitivity was highly and positively correlated with one item only, *annoyance caused by intelligible speech*, ($r_s = .662, p = .010$), while noise sensitivity within a work environment (subscale *work*) was highly correlated with 4 items, namely, *perceived level of global noise* ($r_s = .614, p = .019$), *annoyance caused by perceived global noise* ($r_s = .654, p = .011$), *annoyance caused by intelligible speech* ($r_s = .760, p = .002$), and *distraction caused by noise in the office* ($r_s = .556, p = .039$). Two conclusions can be drawn from these correlations. First, people sensitive to noise were more likely to be annoyed and distracted by the background noise, and they perceived a higher level of noise in their work environment. Second, using the subscale *work* of the noise sensitivity questionnaire seems more appropriate than using the overall score as the former was highly correlated with 4 variables, while the latter was only highly correlated with one variable.

Finally, the background noise level, which included the water sound, was positively correlated with 3 variables, namely, the *possibility to personalise work area* ($r_s = .583, p = .036$), the *satisfaction with Comfort/Functionality subscale of the physical work environment* ($r_s = .570, p = .042$), and the *possibility of working uninterrupted* ($r_s = .652, p = .016$). These positive correlations suggest that the higher the background noise (i.e., the water sound) the more people were satisfied with the comfort/functionality aspects of their work environment, and the more they felt to be able to work uninterrupted, as well as personalise their work area. Furthermore, the background noise was negatively correlated with the *annoyance caused by intelligible speech* ($r_s = -.605, p = .028$), and the *distraction caused by noise in the office* ($r_s = -.744, p = .004$). These correlations are not

surprising, as when the $L_{Aeq,15s}$ was measured at each workstation, the highest $L_{Aeq,15s}$ was 43.6 dB at the closest workstation to the water feature, and the energy averaged $L_{Aeq,15s}$ was 39.3 dB, which is roughly 6 dB lower than the preferred sound pressure level of 45 dBA determined in a previous stage of this study (see Chapter 4) . It appears that the closer the background noise (including the water sound) to the preferred masking level of 45 dBA, the higher the likelihood of making positive changes in the way people perceived their work environment.

7.8 Discussion

This case study provided an insight into using a water feature as a speech masker in an open-plan-office over a period of 3 weeks. Generally, the initial satisfaction in the work environment was high, however, the inclusion of the water feature further increased the satisfaction level of the comfort/functionality aspects of the physical work environment. This increase in satisfaction level was mainly attributed to an increase in people's satisfaction within the noise environment and cleanliness of their work area. The water feature decreased the perceived frequency of occurrence of intelligible speech, which shows its effectiveness in masking speech. This was further demonstrated by an increase in the perceived frequency of occurrence of unintelligible speech. When the water feature was in operation, portions of intelligible speech were masked and, hence converted to unintelligible speech.

The strong and negative correlation between the background noise (including the water sound) and the annoyance caused by intelligible speech further demonstrated the suitability of using water sounds as speech masking sounds. The water sound masked intelligible speech and improved the way people perceived their work environment. The only adverse effect of the water feature was a slight increase in the frequency of going to the toilet reported by two participants out of fourteen. However, the advantages clearly outweigh this small disadvantage.

The small sample size ($N = 14$) is one of the limitations of the study that makes generalisation of the findings difficult beyond the sample tested. Another limitation is that no comparison between male and female participants was possible, as the sample mainly consisted of females ($N=12$). The study was also based on a relatively small and quiet open-plan office, whose people were relatively satisfied before installing the water feature. It is not clear if similar findings would be achieved in more crowded and larger offices with less satisfied people. Having said that, statistically significant improvements

in the work environment were still possible despite the small sample size. In many cases where statistical significance was not found, the magnitude of the effect size was still above .30, which is considered a medium effect size (Cohen, 1988).

More statistically significant results (i.e., $p < .05$) would have been possible had the sample size been larger. In a few cases, the effect sizes of the water feature exceeded .30, but no statistically significant differences were detected. For example, the water feature had a *medium* effect size ($r > .30$) on improving the physical work environment, and on the possibility of concentrating on tasks, but with p -values greater than .05. An effect size of .30 is considered a *medium* effect and could be practically meaningful. For instance, only 1% increase of performance of employees in the office where the water feature was placed, would result in a saving of around £3500 per annum. The calculation was made assuming an average salary of £25000/year and 14 employees ($1\% \times 14 \times 25000$).

From the maintenance point of view, the maintenance required for the water feature was very minimal. The water feature used in this study needed to be filled up once a week, as a portion of the water would evaporate. Additionally, an additive was added to the water to prevent algae build up and keep the water clean and clear. The power consumed by the water pump was 25 W, and the water pump was in operation approximately 8 hours per day, 5 days per week, which results in 1 unit of energy (i.e., 1 kWh) per week. According to statistics published by the Department for Business, Energy and Industrial strategy in the UK, the average price of 1 unit electricity is £0.143. Hence the cost of operating the water feature is only around £0.14 per week, which adds up to roughly £8 per annum (excluding the standing charges and VAT, which is currently at 5%). In other words, operational costs are expected to be completely negligible in comparison to potential savings.

The experiment shows a large potential of using water features in open-plan offices to mask irrelevant speech, making it a very promising topic for further research. Investigating larger, more crowded and busier offices would be the first step towards further research. Furthermore, a longer period (say one year) would be needed to fully understand the benefits and disadvantages of having a water feature in a working area. Investigating multiple water features in one space would be a logical step toward obtaining more conclusive and evidence-based findings.

It is worth pointing out that the water feature remained in the spaces for nearly 11 months after the end of the experiment, because people who were working there wanted to keep

the water feature on a permanent basis. Although no further questionnaires were distributed during the 11 months, the verbal feedback from people working in the space was very positive.

7.9 Conclusions

A cascade-like water feature was placed in a small-sized open-plan office for 3 weeks, and a satisfaction questionnaire was distributed before and after installing the water feature, in order to analyse its impact. Fourteen participants responded to the questionnaire which measured various aspects of the work environment. The results showed that participants were initially moderately satisfied with comfort/functionality, and control/privacy aspects of the physical work environment. Including the water feature in the space significantly improved comfort/functionality aspects of the environment. The improvement in this aspect was attributed to a higher satisfaction level with the global noise environment and the cleanliness of the space.

The water feature significantly reduced the perceived frequency of hearing intelligible speech, whilst it insignificantly increased the perceived frequency of hearing unintelligible speech, which shows the masking capability of the water sound. The tasks that were reported to be affected by intelligible speech were writing, reading, data entry, record updating and telephone conversations. The presence of the water feature did not affect the type of tasks that were reported to have been affected by intelligible speech. Telephone ringing, intelligible speech and unintelligible speech were ranked as highly annoying noise sources, while ventilation/air-conditioning noise was ranked as least annoying noise sources.

The water sound significantly improved the way participants perceived their sound environment. The pleasantness level of the sound environment and the possibility of concentrating on tasks both increased significantly after installing the water feature. Furthermore, the perceived distraction level in the space was significantly reduced by the water sound. Despite a slight increase, no significant differences were detected in the possibility of having a meeting without distracting others, possibility of working uninterrupted for long periods, and possibility of having confidential conversations. The water feature itself was very highly rated as being pleasant, improving the sound environment and being visually pleasing. The water feature was neutrally rated (scores close to 5 out of 10) as helping to carry out private conversations. Only two participants, out of fourteen, reported an increase in the frequency of going to the toilet associated with

the water feature, with an average increase of 2 times per a day. Thirteen participants (out of 14) preferred the water feature to remain in the space on a permanent basis. The comments in the open-ended question were all positive; e.g., “very pleasant”, “it is a pleasant soothing sound”, “the feature is lovely to look at”. The responses obtained in the questionnaire suggest that the water feature was highly appraised by participants with very little adverse effect on the number of times that people needed to go to the toilet.

Correlation analysis revealed that the age of participants positively correlated with the *physical position of work station*, the *possibility of not being seen by others*, and the *possibility of meeting without distracting others*. Whilst the age of participants negatively correlated with the *perceived frequency of hearing unintelligible speech*. Overall, noise sensitivity scores did not seem to strongly correlate with the variables, except for the annoyance caused by intelligible speech, where a strong and positive correlation was detected. However, the noise sensitivity score within the subscale *work* strongly correlated with four variables, namely, the *perceived level of global noise*, the *annoyance caused by perceived global noise*, the *annoyance caused by intelligible speech*, and, the *distraction caused by noise in the office*. This suggests that scores from the subscale *work* might be more appropriate to be used in studies related to open-plan offices, in comparison to the overall noise sensitivity score. Background noise in the office (which included the water sound) positively correlated with the *possibility to personalise work area*, the satisfaction with *Comfort/Functionality* subscale of the physical work environment, and the *Possibility of working uninterrupted*. It is interesting that the background noise of the space had a negative correlation with the *annoyance caused by intelligible speech*, which once again shows the masking ability of the water sound. It is worth mentioning that despite all the positive responses and effect of the water feature on the work environment, the findings cannot be generalised beyond the sample studied, due to the small sample size. However, the likelihood of achieving similar responses in larger samples is very high, given the statistically medium to large effect sizes reported throughout this study.

In the next chapter, objective measures are provided regarding the reduction in distraction and privacy distances associated with adding a water sound to an open-plan office, to relate subjective responses reported in this chapter to objective measures.

CHAPTER 8: OBJECTIVE MEASURES

8.1 Introduction

In previous chapters, numerous subjective and objective experiments were carried out that involved human participation. The current chapter further extends previous findings through using objective measures to assess the likely improvement in the room acoustic parameters of open-plan offices, associated with using a water sound as a speech masking sound. The chapter aims at providing a practical guideline on what should be expected from installing a water feature in an open-plan office in view of reducing the distraction and privacy distances which are associated with the reduction of the speech transmission index (STI) in the space. The distraction and privacy distances are quantities used alongside other measures to assess room acoustic parameters in open-plan offices. These measures are known as single-number quantities. To meet the aim of this part of the study, these single-number quantities were measured in two open-plan offices before and after including a water sound in the sound environment of those spaces. A brief description of the single-number quantities is initially given, followed by an in-depth explanation of the methodologies used to measure them. Results obtained from these measurements are then presented and recommendations are given on how the results should be interpreted. The chapter ends with a critical discussion of the findings obtained, followed by the conclusions of the chapter.

8.2 Single-number quantities

Virjonen *et al.* (2009) proposed single-number quantities that were later adopted in BS EN ISO 3382-3 (2012) to provide a standardized method representing room acoustic parameters in open-plan offices. These single-number quantities and their history were presented in Chapter 2 and Chapter 3. A brief description of the single-number quantities is given here. The quantities are the spatial decay rate of the A-weighted sound pressure level (SPL) of speech, $D_{2,S}$, the A-weighted SPL of speech at 4 metres, $L_{p,A,S,4m}$, and the distraction distance, r_D . The BS EN ISO 3382-3 (2012) also recommends measuring the STI in the nearest workstation, the averaged A-weighted background noise, $L_{p,A,B}$, and the privacy distance, r_P . The procedures involved in measuring these single number quantities are explained in this chapter, and measured in accordance with appropriate standards.

8.3 Water sound

In line with the preference and perception tests, the 37-jet fountain sound (FTW) was chosen to be used in this study as a speech masking sound. FTW was most preferred in the audio-only preference test and was highly rated in the perception test (Chapter 5). Acoustic and psychoacoustic characteristics of the water sound was given in Chapter 3 and Chapter 5.

8.4 Equipment and setup

Two sets of equipment were used to carry out the measurements. The list of equipment used in the experiment is presented in Table 8.1. One set of equipment was used to measure the attenuation of sound pressure level in the spaces and the impulse response, from which the STIs were derived. The other set was used to measure the background noise before and after adding the water sound.

The Maximum-Length Sequence System Analyzer (MLSSA) software, installed on a personal computer, was used to measure the impulse response and compute the modulation transfer function, MTF, from which the STI was calculated. The signal generated by MLSSA was amplified using the power amplifier Brüel and Kjær Type 2706 and equalised using a stereo graphic equaliser (Technics Model NO. SH-8075) before it was sent to a custom-made omnidirectional loudspeaker. The equaliser was used to finetune the output signal to meet the requirements set out by BS EN ISO 3382-1 (2009) for omnidirectional sound sources. The omnidirectional loudspeaker was placed on a tripod 1.2 m above the ground from the centre of the loudspeaker, and was used to play the signal generated by the MLSSA, which was effectively a pink noise. The signal (i.e., the pink noise) was then captured by a microphone (Brüel and Kjær Type 4190), placed on a tripod 1.2 m above the ground. The microphone was powered by a microphone power supply (Brüel and Kjær Type 2804) and was connected to a sound measuring system (Norsonic 823) which acted as an amplifier amplifying the signals captured by the microphone before feeding them back into MLSSA to derive the MTFs. The same signal was used to measure the attenuation in sound pressure level as a function of distance from the sound source. The water sound was played from a laptop, Dell XPS L502X, which was connected to an external USB soundcard (M-Audio MobilePre). The soundcard was connected to an amplifier (J.E.Sugden A28), and the amplifier was connected to a loudspeaker (KEF Coda III Type SP 3016). The loudspeaker was placed on a table

making the acoustic centre of the loudspeaker 1 m above the ground. Figure 8.1 shows a schematic diagram of the equipment configurations.

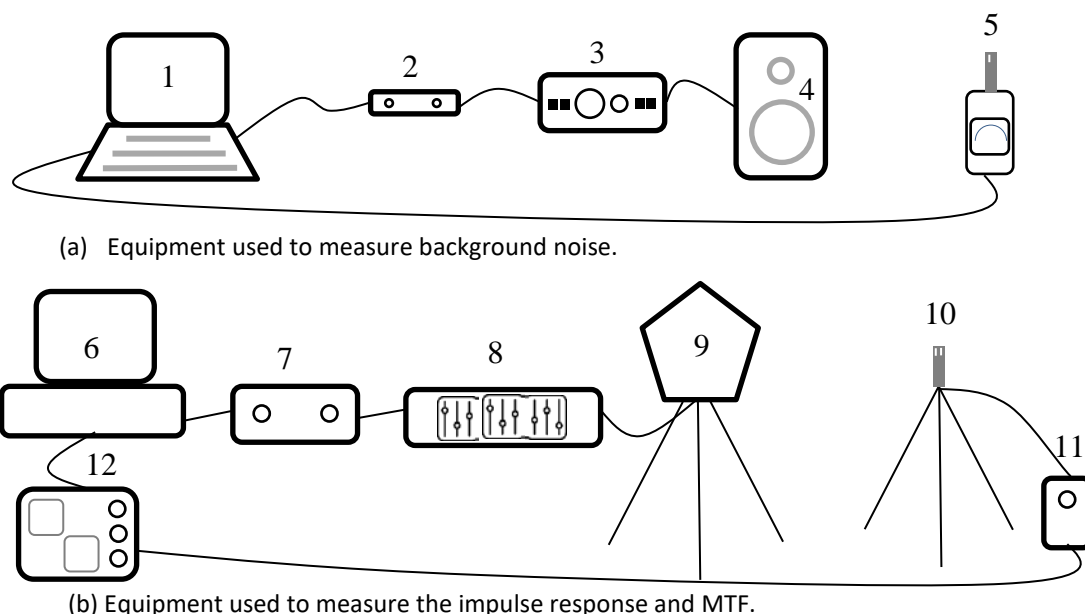


Figure 8.1 Equipment used to measure the single-number quantities. Descriptions of the equipment are given in Table 8.1.

Table 8.1 List of equipment used to measure the single-number quantities and their descriptions.

#	Type	Model	Purpose
1	Laptop	Dell XPS L502X	Digital audio processing.
2	Soundcard	M-Audio MobilePre USB	Control the sound level of audio signals.
3	Power amplifier	J.E.Sugden A28	Amplify signals sent from the soundcard to the loudspeaker.
4	Loudspeaker	KEF Coda III Type SP 3016	Play the water sound.
5	Hand-held analyser	Brüel and Kjær Type 2250	Measure octave-band sound pressure level of the background noise with and without water sound.
6	Personal computer	Viglen Contender	Operate MLSSA.
7	Power amplifier	Brüel and Kjær Type 2706	Amplify audio signals sent from the personal computer to the omnidirectional speaker.
8	Stereo Graphic Equaliser	Technics Model NO. SH-8075	Fine tune audio signal to generate a pink noise whose properties meet requirements set out by EN ISO 3382-1 (2009).
9	Omnidirectional loudspeaker	Custom-made	Play pink noise signal sent from the personal computer.
10	Microphone	Brüel and Kjær Type 4190	Record impulse response and sound attenuation at the receiver position.
11	Microphone power supply	Brüel and Kjær Type 2804	Supply the microphone with power to operate.
12	Sound measuring system	Norsonic 823	Amplify impulse response captured by the microphone so that the MLSSA can analyse it and derive the NTFs.

8.5 Measurement procedure

Measurements in this part of the research were carried out in accordance with BS EN ISO 3382-3 (2012). The open-plan offices used in this experiment were furnished, and the measurements were carried out in the absence of people, except the researcher who carried out the measurements. This represents the worst-case scenario, as in the presence of people, the background noise level in the space increases as a result of noise generated from work related activities, which in turn reduces the signal-to-noise ratio and ultimately the speech intelligibility.

The omnidirectional loudspeaker was placed at one of the workstations, 1.2 m above the ground, and used as a speech source. Measurements were carried out along a line which crossed over workstations starting from the workstation closest to the source, to the farthest workstation. However, only measurements within the range 2 m to 16 m away from the sound source were used for the determination of $D_{2,s}$, as per recommendations by BS EN ISO 3382-3 (2012). The number of microphone positions used for each source varied from 6 to 11 depending on the sizes of the open-plan offices. The minimum number of microphone positions required by BS EN ISO 3382-3 (2012) is 4. For each room, two sound source positions were used. The source positions were determined based on the layout of the room tested. The single number quantities were calculated for each sound source separately. The sound pressure level of the water sound was set at 48 dBA at the closest workstation from the loudspeaker. This is in line with finding achieved from the sound level preference test, in which 45 dBA was most preferred, closely followed by 48 and 42 dBA. The preference scores given to the level 48 and 42 were not significantly different from that of level 45 dBA. Setting the SPL of the water sound at 48 dBA at the closest workstation allowed for the majority of workstations to have an SPL close to 45, and the farthest to have levels close to 42 dBA.

At each measurement point (i.e., microphone position), four measurements were carried out:

1. Sound pressure level in octave bands (125 to 8000 Hz) of pink noise, $L_{p,Ls}$.
2. STI.
3. Background noise level in octave bands, $L_{p,B}$.
4. Distance to the sound source, r .

The water sound was added to the spaces through a high quality loudspeaker which was placed on a stand, making the acoustic centre of the loudspeaker 1 m high from the ground. The position of the loudspeakers was carefully chosen in both spaces, to ensure a uniformly distributed sound level in the spaces.

8.5.1 Determination of the sound power level of an omnidirectional speaker

The sound power level of the omnidirectional speaker, $L_{w,LS}$, was determined in accordance with BS EN ISO 3741 (2010). The measurement was carried out in the reverberant chamber of the acoustic laboratory at Heriot-Watt University, using the equivalent sound absorption area of the reverberation test room method, commonly known as the *direct method*. The loudspeaker was placed on a tripod 1.2 m above the ground and at least 1.5 m from other boundaries of the room. Six initial microphone positions were selected at which the sound pressure level of the loudspeaker was measured. Microphone positions met the requirements set out in Section 8.3 of BS EN ISO 3741 (2010). The standard deviation, s_M , of the measured SPLs at six initial microphone positions for each one-third octave band (from 125 to 8000 Hz) was calculated using Equation (8.1)

$$s_M = \sqrt{\sum_{i=1}^{N_{M(pre)}} \frac{[L'_{pi(pre)} - L'_{pm(pre)}]^2}{N_{M(pre)} - 1}} \quad (8.6)$$

where

- $L'_{pi(pre)}$ is the one-third-octave band time-averaged sound pressure level measured at the i th initial microphone position, with the noise source under test in operation, in decibels;
- $L'_{pm(pre)}$ is the arithmetical mean value of the one-third-octave band time-averaged sound pressure levels measured at the six initial microphone positions, with the noise source under test in operation, in decibels;
- $N_{M(pre)}$ = 6, the initial number of microphone positions.

The standard deviation of the sound pressure levels, s_M , obtained from Equation (8.1) did not exceed 1.5 dB, and therefore, no additional microphone/source positions were needed and the six initial microphone positions were taken as final measurements, $L_{pi(ST)}$. At each microphone position, the sound pressure level in each one-third octave band was at least

15 dB higher than the background noise, and thus no correction for the effect of background noise was applied to the measurements.

The mean time-averaged sound pressure level in the test room with the noise source under the test in operation, $\overline{L_{p(ST)}}$, was calculated using Equation (8.2):

$$\overline{L_{p(ST)}} = 10 \log \left[\frac{1}{N_M} \sum_{i=1}^{N_M} 10^{0.1 L_{pi(ST)}} \right] \text{ dB} \quad (8.7)$$

where

- $L_{pi(ST)}$ is the one-third-octave band time-averaged sound pressure level at the i th microphone position, with the noise source under the test in operation, in decibels;
- N_M is the number of microphone positions (i.e., 6).

The sound power level of the omnidirectional loudspeaker in each one-third octave band, $L_{W,LS,i}$, under reference meteorological conditions, was calculated using Equation (8.3):

$$L_{W,LS,i} = \overline{L_{p(ST)}} + \left\{ 10 \log \frac{A}{A_0} + 4.34 \frac{A}{S} + 10 \log \left(1 + \frac{S c}{8 V f} \right) + C_1 + C_2 - 6 \right\} \text{ dB} \quad (8.8)$$

where

- $\overline{L_{p(ST)}}$ is the mean one-third-octave band time-averaged sound pressure level in the test room with the omnidirectional loudspeaker in operation in decibels;
- A is the equivalent absorption area, in square metres, of the room:

$$A = \frac{55.26}{c} \left(\frac{V}{T_{60}} \right),$$
in which T_{60} is the reverberation time, in seconds, of the reverberant test room at mid-band frequency of the measurement;
- A_0 = 1 m²;
- S is the total surface area, in square metres, of the reverberation test room;
- c is the speed, in metres per second, of sound at temperature, θ , in degrees Celsius, of the air in the reverberation test room at the time of test:

$$c = 20.05 \sqrt{273 + \theta};$$

V	is the volume, in cubic metres, of the reverberation test room;
f	is the mid-band frequency, in hertz, of the measurements;
C_1	is the reference quantity correction, in decibels, to account for the different reference quantities used to calculate decibel sound pressure level and decibel sound power level, and is a function of the characteristic impedance of the air under the meteorological conditions at the time and place of the measurements: $C_1 = -10 \log \frac{p_s}{p_{s,0}} + 5 \log \left(\frac{273.15 + \theta}{\theta_0} \right);$
C_2	is the radiation impedance correction, in decibels, to change the actual sound power relevant for the meteorological conditions at the time and place of the measurement into the sound power under reference meteorological conditions: $C_2 = -10 \log \frac{p_s}{p_{s,0}} + 15 \log \left(\frac{273.15 + \theta}{\theta_1} \right);$
p_s	is the static pressure, in kilopascals, in the test room at the time of test;
$p_{s,0}$	is the reference static pressure, 101.325 kPa;
θ	is the air temperature, in degrees Celsius, in the test room at the time of test;
θ_0	= 314 K; and
θ_1	= 296 K.

The reverberation time, T_{60} , in the test room (i.e., the reverberant chamber), was measured in accordance with BS EN ISO 3382-2 (2008), except that only the first 15 dB decay, T_{15} was used. Three source positions, each with 5 microphone positions were used to measure the reverberation time.

8.5.2 Determination of the spatial decay

After the determination of the sound power level of the omnidirectional source, the first measurement in the open-plan office was to measure the noise-free impulse response and calculate the MTFs. The second step was to measure the attenuation in sound pressure level as a function of distance from the noise source. In the third and last step, the background noise (with and without the water sound) at each microphone position was measured. Data acquired within the second and third steps were used to correct the noise-free MTFs to take into account the effects of background noise and speech level.

The sound power spectrum of normal speech having octave band values of normal effort unisex speech (average of male and female speech) was used. The octave band sound power levels as well as the sound pressure level at 1 m away from the acoustic centre of the sound source in the free field ($L_{p,S,1\text{ m}}$) are given in Table 8.2. The resulting A-weighted sound power level and sound pressure level are 68.4 dBA re 10^{-12} W and 57.4 dBA respectively.

Table 8.2 The sound levels of normal effort speech at distance of 1 m in free field from the speaker and the A-weighting of octave bands (BS EN ISO 3382-3, 2012).

Band No. i	Frequency Hz	Sound power level $L_{W,S}$ dB re 10^{-12} W	Sound pressure level $L_{p,S,1\text{ m}}$ dB re 2×10^{-5}	A-weighting dB
1	125	60.9	49.9	-16.1
2	250	65.3	54.3	-8.6
3	500	69.0	58.0	-3.2
4	1000	63.0	52.0	0.0
5	2000	55.8	44.8	1.2
6	4000	49.8	38.8	1.0
7	8000	44.5	33.5	-1.1
A-weighted		68.4	57.4	

The sound power level and the sound pressure level presented in Table 8.2 were only used in the post-processing of measurements. The actual output of the omnidirectional loudspeaker was a pink noise that was sufficiently higher than the background noise (i.e., at least 20 dB) at the most distant measurement point. The omnidirectional loudspeaker was placed in the open-plan offices at the selected source position then the sound pressure level in the n selected microphone position, $L_{p,LS,n,i}$, was recorded. The measurement was corrected to eliminate the contribution of background noise in accordance with ISO 3741 (2010). The sound power level of the omnidirectional speaker calculated in the previous section, $L_{W,LS,i}$, was then converted into sound pressure level at 1 m, $L_{p,S,1\text{ m},i}$, using Equation (8.4):

$$L_{p,S,1\text{ m},i} = L_{W,LS,i} + 10 \log \frac{1}{4\pi \times 1.0^2} \approx L_{W,LS,i} - 11 \text{ dB} \quad (8.9)$$

where

- $L_{p,S,1\text{ m},i}$ is the sound pressure level at 1 m, in octave bands (dB);
- $L_{W,LS,i}$ is the sound power level of the loudspeaker in octave bands (dB);
- i denotes the octave band.

The difference in dB between the sound pressure level of pink noise at 1 m from the source, $L_{p,S,1\text{ m}}$, and the measured sound pressure level at the measurement point n at distance r_n , is called the attenuation $D_{n,i}$, and is determined by

$$D_{n,i} = L_{p,S,1\text{ m},i} - L_{p,LS,n,i} \quad (8.10)$$

where

- $L_{p,S,1\text{ m},i}$ is the sound pressure level at a distance of 1 m (dB);
- $L_{p,LS,n,i}$ is the sound pressure level at the measurement point n (dB);
- i denotes the octave band.

Attenuation is independent on the sound pressure level and once found, it is applicable to any kind of sound spectra. Hence, the sound pressure level of normal speech at position n and octave band i , $L_{p,LS,n,i}$, is the sound pressure level of normal speech (Table 8.2) deducted by $D_{n,i}$:

$$L_{p,S,n,i} = L_{p,S,1\text{ m},i} - D_{n,i} \quad (8.11)$$

where

- $L_{p,S,1\text{ m},i}$ is the sound pressure level of normal speech at a distance of 1 m from the omnidirectional source (dB);
- $D_{n,i}$ is attenuation at measurement point n (dB);
- i denotes the octave band.

The A-weighted speech level at position n , $L_{p,A,S,n}$, was finally obtained by adding the value of A-weighting factors at each octave band and summing them up on an energy basis as follows

$$L_{p,A,S,n} = 10 \log \left(\sum_{i=1}^7 10^{0.1(L_{p,S,n,i} + A_i)} \right) \quad (8.12)$$

where

- $L_{p,S,n,i}$ is the sound pressure level of normal speech at the measurement point n obtained from Equation (8.6);
- A_i is the A-weighting correction presented in Table 8.2;
- i denotes the octave band.

The spatial decay of A-weighted speech $D_{2,S}$ is then determined from the results at measurement positions using the least square method as well as a logarithmic distance axis and linear regression as follows

$$D_{2,S} = -\log_2 \left| \frac{N \sum_{n=1}^N \left[L_{p,A,S,n} \log \left(\frac{r_n}{r_0} \right) \right] - \sum_{n=1}^N L_{p,A,S,n} \sum_{n=1}^N \log \left(\frac{r_n}{r_0} \right)}{N \sum_{n=1}^N \left[\log \left(\frac{r_n}{r_0} \right) \right]^2 - \left[\sum_{n=1}^N \log \left(\frac{r_n}{r_0} \right) \right]^2} \right| \quad (8.13)$$

where

- $L_{p,A,S,n}$ is the A-weighted speech level in position n (dBA);
- n is the index number of the single measurement position;
- N is the total number of measurement positions;
- r_n is the distance to measurement position n (m);
- r_0 is the reference distance, 1 m.

Only results measured within the range of 2 to 16 m from the acoustic source are included within the calculation process. Once the $D_{2,S}$ was known, the A-weighted SPL of speech at 4 m, $L_{p,A,S,4\text{ m}}$ was determined using a linear regression line determined from the A-weighted speech level in all microphone positions, $L_{p,A,S,n}$, as a function of distance on a logarithmic axis.

8.5.3 Determination of distraction distance, r_D and privacy distance, r_P

The distraction distance, r_D , is the distance at which STI drops below a value of 0.50, and the privacy distance, r_P , is the distance at which the STI drops to a value of 0.20. At each microphone position, the STI was calculated twice; one without the effect of the water sound and one with the effect of the water sound. The indirect STI method using the impulse response and forward energy integral (Schroeder integral) was used to derive the modulation transfer functions in accordance with IEC 60268-16 (2011). The Maximum-Length Sequence System Analyzer program (MLSSA), installed on a personal computer, was used to measure the impulse response and compute the modulation transfer functions, MTFs, from which the STI was calculated. A noise-free impulse response (i.e., SNR > 20 dB) was measured and corrected for the effects of the background noise and speech level by post-processing, as described later. Hence, the noise-free impulse response was measured once, from which the noise-free MTFs were calculated. The noise-free MTFs were corrected for the effect of speech level by post-processing using the normal speech spectrum given in Table 8.2 and applying the attenuation, $D_{n,i}$, that was calculated in the previous section using Equation (8.5). The effect of background noise and speech level

was added to the STI calculation by measuring the sound pressure level of the background noise, $L_{p,B}$, at each microphone position in octave bands, with and without the water sound. The A-weighted sound pressure level, $L_{p,A,B}$ was determined accordingly, and the average background noise level of all microphone positions was calculated and used to determine the corrected MTFs at each microphone position using Equation (8.9) (IEC 60268-16, 2011):¹

$$\text{MTF}(f_m) = \frac{|\int_0^\infty h_k(t) e^{-j2\pi f_m t} dt|}{\int_0^\infty h_k(t)^2 dt} \times [1 + 10^{-SNR_k/10}]^{-1} \quad (8.14)$$

where

- $h_k(t)$ is impulse response of octave band k ;
- f_m is the modulation frequency;
- SNR_k is the signal-to-noise ratio in dB.

The first factor of Equation (8.9) is Schroeder's equation which was obtained from the MLSSA program. The second factor of Equation (8.9) is the background noise speech level effect which was added to the measured MTFs from MLSSA by post-processing. When the SNR is higher than 20 dB the second factor of Equation (8.9) approaches 1 and the resulting MTF is considered to be noise free. The pink noise used by MLSSA to generate the MTFs, had a significantly higher sound pressure level (≥ 20 dB) than the background noise and therefore, it is safe to state that the MTFs were noise-free.

Once the MTFs were corrected for the effect of speech level and background noise (using Microsoft Excel), the STI was calculated at each microphone position. The process of calculating the STI from MTFs was explained in detail in Chapter 2. The distraction distance, r_D and privacy distance, r_P were calculated using a linear regression line determined from the STI values as a function of the distance from the source position on a linear axis.

Both $D_{2,S}$ and $L_{p,A,S,4m}$ are not affected by the background noise of the space. Therefore, the addition of the water sounds would not change their values. However, they were included in the experiment to put the change in r_D and r_P into context and allow a more accurate interpretation of these changes as a function of the room acoustics of the spaces.

¹ The term MTF (f_m) is used in Equation (8.9) for keeping consistency across the chapter. The actual term used in the IEC 60268-16, (2011) is $m_k(f_m)$.

8.6 Open-plan offices

Two open-plan offices were tested. These differed in size, layout and finishing materials. The first open-plan office (Office 1) was the Computer room of the Resource Centre in the Edwin Chadwick Building at Heriot-Watt University. A photograph of the space is provided in Figure 8.2. Dimensions of the room were 7.28 m wide \times 14.20 m deep \times 2.95 m high. Figure 8.4 shows a plan of the space with its dimensions. The space was a computer room for students which accommodated 33 computer desks clustered into four working zones, as shown in Figure 8.4. No partition screens were installed between the workstations. The walls were made of painted concrete blocks, and the ceiling was made of absorptive tiles. The floor was covered by a 5-mm-thick carpet. Windows accounted for 7% of the total area of the walls. The space was equipped with an air-conditioning system, but it was not in operation during the measurements due to maintenance work.

The second open-plan office (Office 2) was the ground floor of the William Arrol Annexe, at Heriot-Watt University. A photograph of the space is provided in Figure 8.3. Dimensions of the room were 7.00 m wide \times 26.05 m deep \times 2.75 m high. Figure 8.5 shows a plan of the space with its dimensions. The space consisted of a modular open-plan office for postgraduate research students and accommodated 44 workstations clustered into 13 working zones. The workstations were separated with 0.4-m-high upholstered partition screens. Due to the office being a modular building, all finishing materials were acoustically reflective. The walls were finished with PVC laminated gypsum boards, the ceiling was covered with PVC tiles, and the finishing material of the floor was linoleum flooring. Windows accounted for 15% of the total area of the wall. The space was equipped with an air-conditioning system.

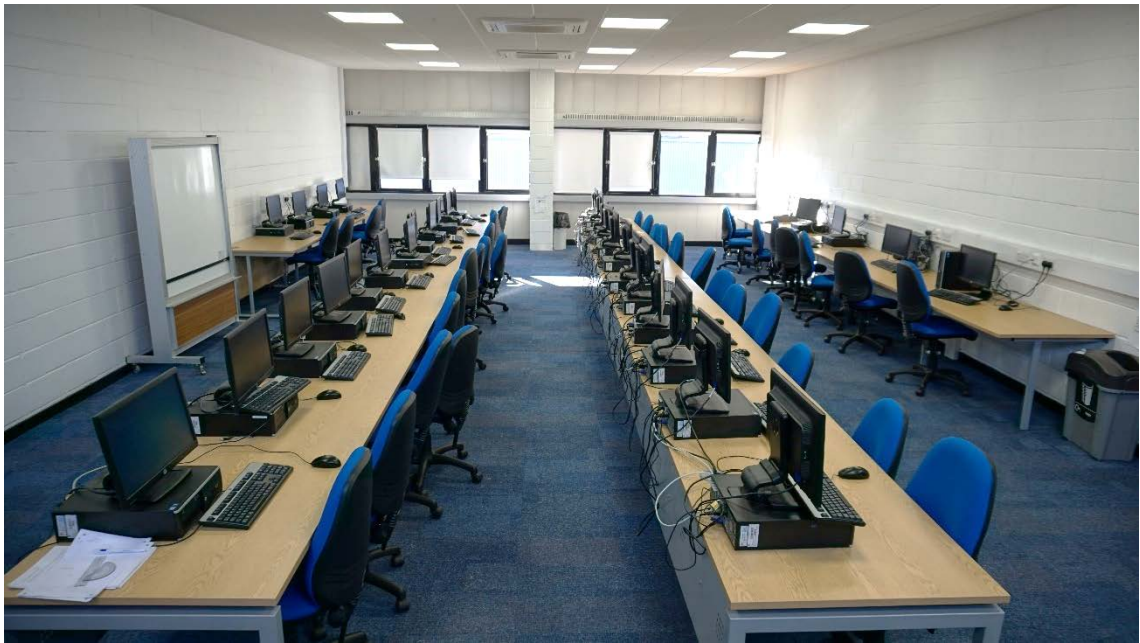


Figure 8.2 A photograph of Office 1. Computer room of the Resource Centre, Heriot-Watt University

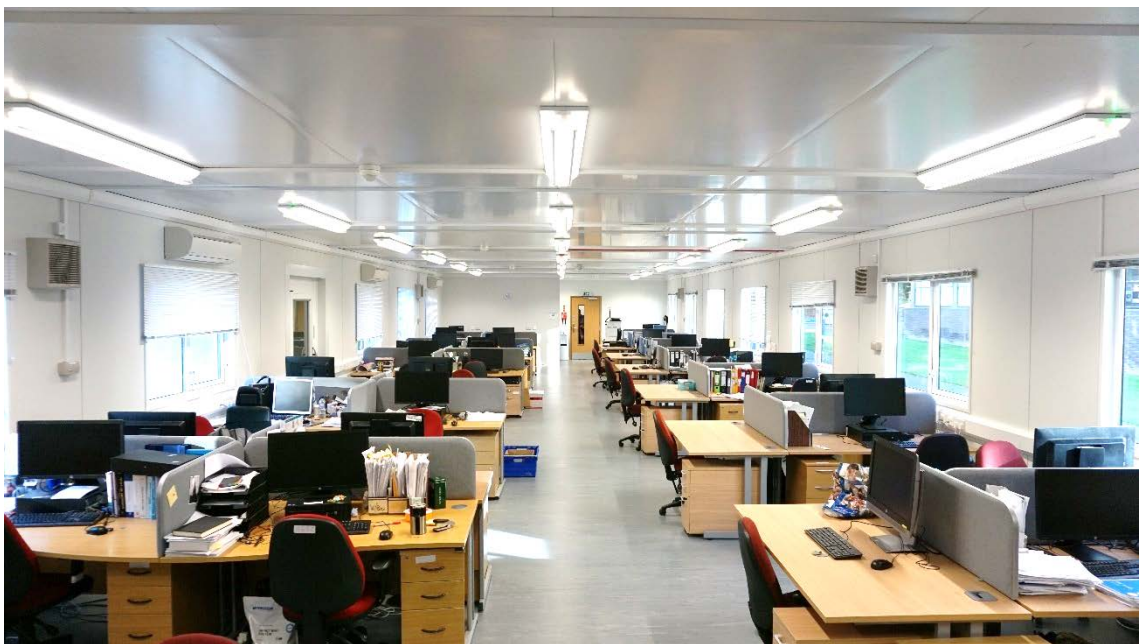


Figure 8.3 A photograph of Office 2. Ground floor of the William-Arrol Annex, Heriot-Watt University

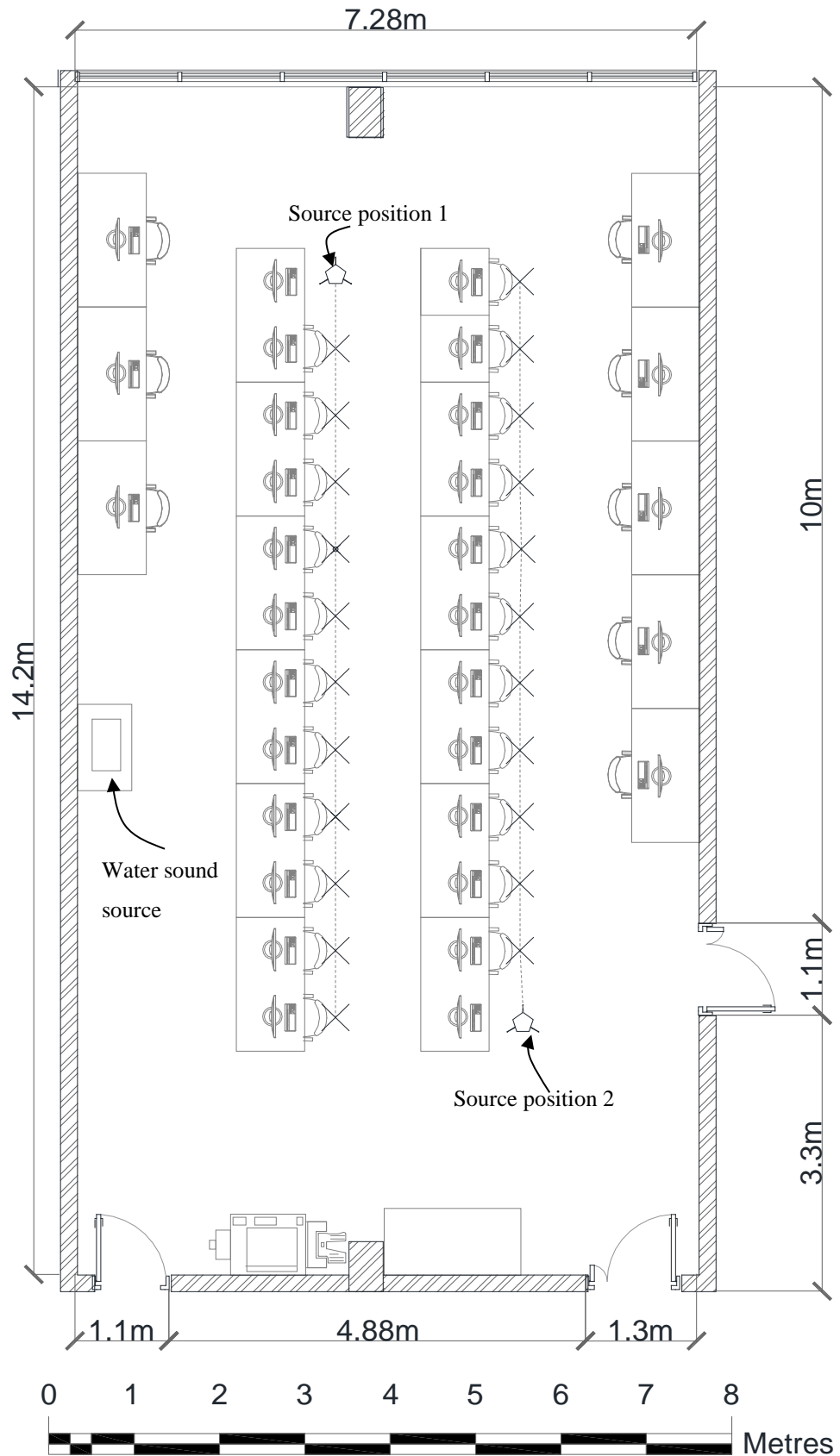


Figure 8.4 Plan and dimensions of Office 1. Dimensions are in metres. Microphone positions are denoted by \times .

8.7 Results

The single-number quantities of the two spaces are presented in this section. In Office 1, r_D and r_P were measured under two masking conditions; no masking, and water sound as a speech masking sound. In Office 2, due to having an air-conditioning system installed in the space, r_D and r_P were measured under four masking conditions, namely, no masking, air-conditioning noise alone, air conditioning noise with the water sound, and water sound alone. Only the values of r_D and r_P are affected by the masking sound, and the remaining quantities are related to the room acoustics of the spaces. Therefore, they were measured once. In both office spaces, the STI value did not fall below 0.20, thus, the privacy distance, r_P , could not be measured directly, and extrapolation was used to estimate its value.

8.7.1 Office 1

The single number quantities measured in Office 1 from two source positions are listed in Table 8.3. The results suggest poor office acoustics in the space. The values of $D_{2,s}$ were -3.4 dB and -3.5 dB for source positions 1 and 2, respectively. This means that the sound pressure level of speech would reduce by around 3.5 dB per doubling of the source-receiver distance in this space. This is understandable as there were no partition screens between workstations and the walls were highly reflective. The A-weighted SPL of speech at 4 metres, $L_{p,A,S,4m}$, was measured to be 53.13 dB and 53.28 dB, for source position 1 and 2, respectively. The spatial decay rate of the A-weighted sound pressure level of speech in Office 1 is shown in Figure 8.6 and Figure 8.7, for source position 1 and source position 2, respectively. The figures show the A-weighted sound pressure level of speech alongside the A-weighted sound pressure level of the background noise at each microphone position. Logarithmic regression was used to generate the trend lines which in turn were used to calculate $L_{p,A,S,4m}$.

The rate of drop in STI as a function of distance from the source, before and after adding the water sound is shown in Figure 8.8 and Figure 8.9, for source positions 1 and 2 respectively. Linear regression was used to generate the trend lines from which the values of r_D and r_P were calculated. The values of r_D , without the water sound, were 14.64 m and 16.04 m, from source position 1 and 2, respectively. After adding the water sound to the background noise, the values dropped to 5.50 m for source position 1 (closer to the water sound source) and 7.40 m for source position 2.

The energy averaged A-weighted background noise in the space was 33.60 dB 34.63 dB for source position 1 and 2 respectively. After adding the water sound, the values increased to 45 dB for source position 1 (microphone positions closer to the water sound) and 42.24 for source positions 2, both of which are within the range of recommended masking levels discussed in Chapter 4.

Table 8.3 Single-number quantities measured in Office 1 from two source positions.

Source position 1 (11 microphone positions)		
Masking condition	NM	WA
STI at the nearest workstation	0.86	0.75
Distraction distance, r_D , in m	14.64	5.50
Privacy distance, r_P , in m (extrapolated)	28.09	14.20
Spatial decay rate of A-weighted SPL of speech, $D_{2,S}$, in dB	-3.40	-3.40
A-weighted SPL of speech at 4 metres, $L_{p,A,S,4\text{ m}}$, in dB	53.13	53.13
Average A-weighted background noise level, $L_{p,A,B}$, in dB	33.60	45.00
Source position 2 (10 microphone positions)		
Masking condition	NM	WA
STI at the nearest workstation	0.85	0.78
Distraction distance, r_D , in m	16.04	7.40
Privacy distance, r_P , in m (extrapolated)	32.00	7.40
Spatial decay rate of A-weighted SPL of speech, $D_{2,S}$, in dB	-3.50	-3.50
A-weighted SPL of speech at 4 metres, $L_{p,A,S,4\text{ m}}$, in dB	53.28	53.28
Average A-weighted background noise level, $L_{p,A,B}$, in dB	34.63	42.24

NM: No masking. WA: Water sound

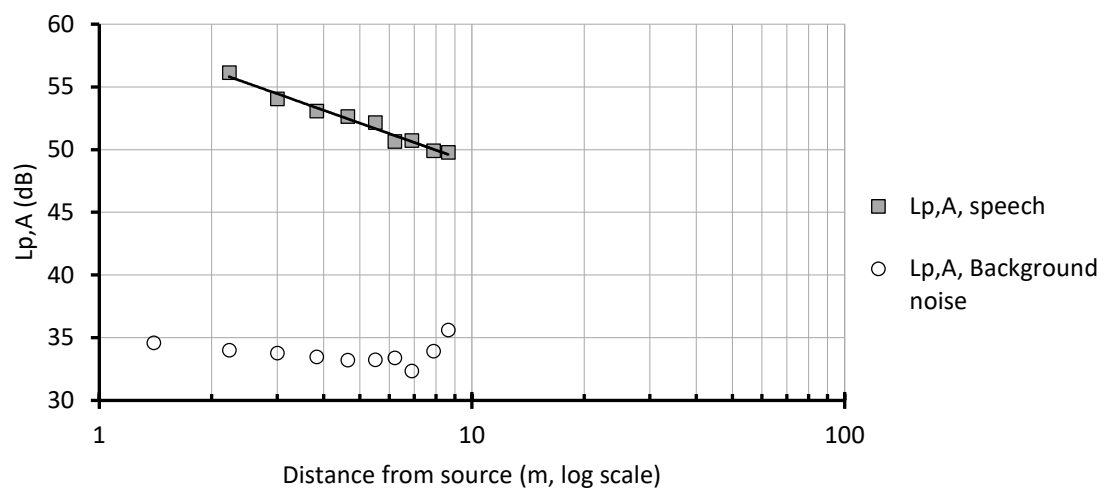


Figure 8.6 Spatial decay rate of the A-weighted sound pressure level of speech in Office 1 from source position 1. Data markers: measured data; trend line: logarithmic regression line.

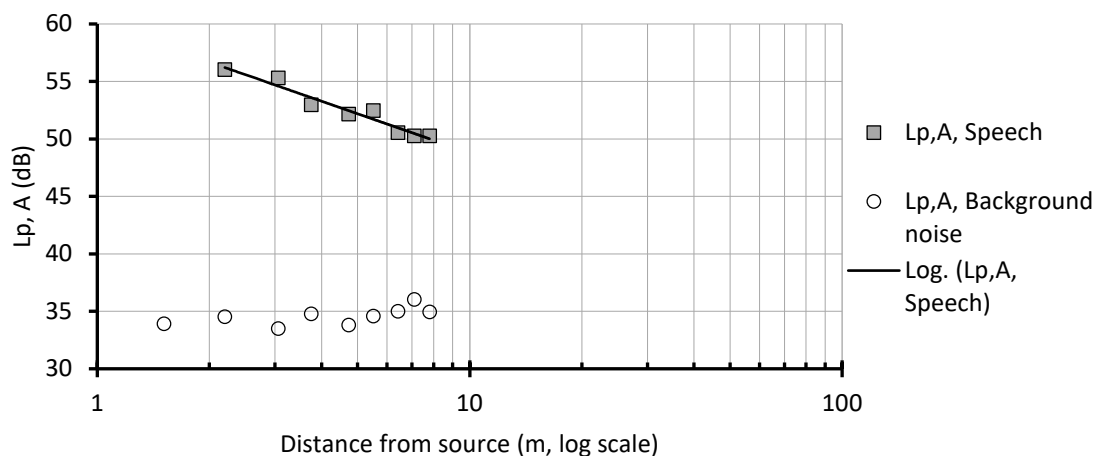


Figure 8.7 Spatial Decay Rate of the A-weighted sound pressure level of speech in Office 1 source position 2. Data markers: measured data; trend line: logarithmic regression line.

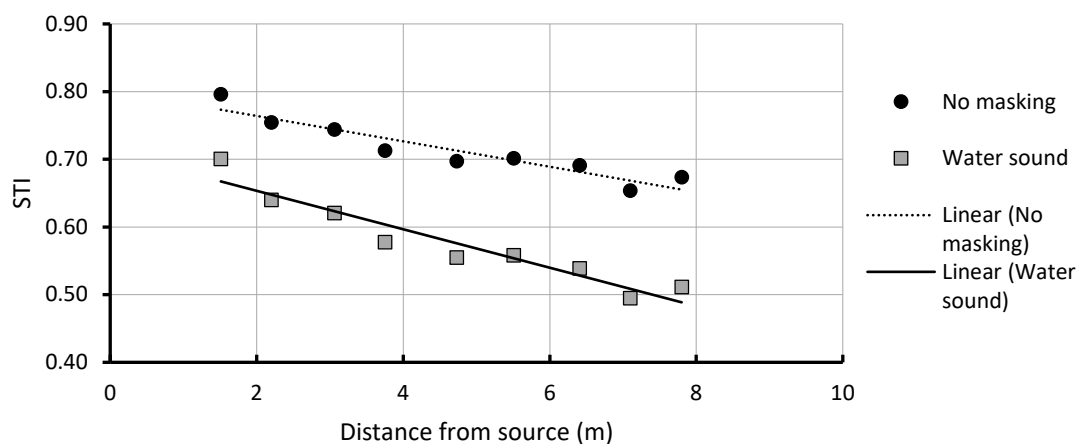


Figure 8.8 Rate of drop in STI in Office 1 measured for source position 1, in two masking conditions: no masking and water sound. Data markers: measured data; trend lines: linear regression line.

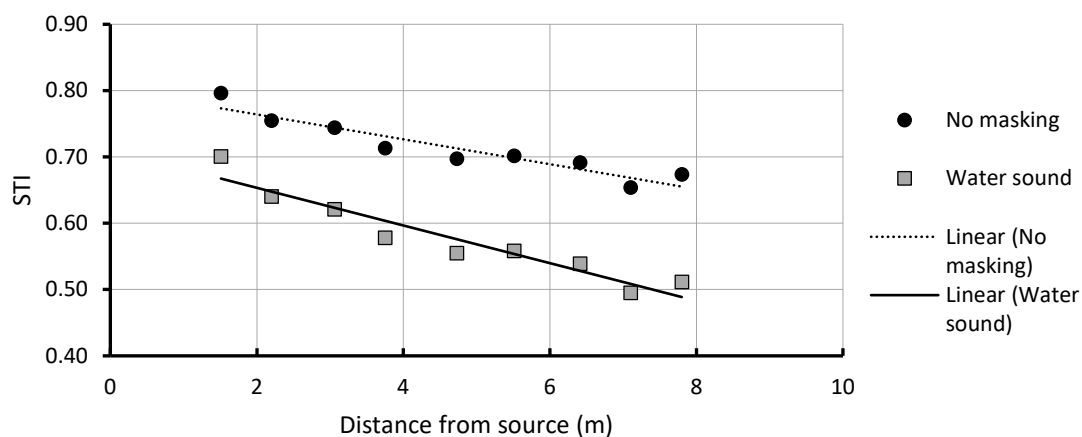


Figure 8.9 Rate of drop in STI in Office 1 measured for source position 2, in two masking conditions: no masking and water sound. Data markers: measured data; trend lines: linear regression line.

8.7.2 Office 2

The single-number quantities measured in Office 2 from two source positions are listed in Table 8.4. As a result of having the air-conditioning system in this office, the r_D and r_P were measured under four masking conditions. The masking conditions were no masking (NM), air-condition noise (AI), air-condition noise with water sound (AW), and water sound (WA). Similar to Office 1, The results suggest poor office acoustics in the space.

The values of $D_{2,S}$ were -3.59 dB and -3.42 dB for source position 1 and 2, respectively. This means the sound pressure level of speech would reduce by only around 3.5 dB per doubling the source/receiver distance in this space. The A-weighted SPL of speech at 4 metres, $L_{p,A,S,4\text{ m}}$, was measured to be 53.70 dB and 53.90 dB, for source positions 1 and 2, respectively. These values are close to those recorded in Office 1 even though Office 2 had partition screens installed between workstations. The reason could be the highly reflective materials used in Office 2, that increased the reverberant field in the space. The spatial decay rate of the A-weighted sound pressure level of speech in Office 2 is shown in Figure 8.10 and Figure 8.11, for source position 1 and source position 2, respectively. The figures show the A-weighted sound pressure level of speech alongside the A-weighted sound pressure level of the background noise at each microphone position. Logarithmic regression was used to generate the trend lines which in turn were used to calculate $D_{2,S}$, and $L_{p,A,S,4\text{ m}}$.

The rates of drop in STI as a function of distance from the source, in all four masking conditions, are shown in Figure 8.12 and Figure 8.13, for source position 1 and 2, respectively. Linear regression was used to generate the trend lines from which the values of r_D and r_P were calculated. As expected, the no masking condition resulted in the greatest value of r_D , which was 13.24 m for source position 1, and 14.31 m for source position 2. Air-conditioning noise resulted in the second highest r_D value, which was 7.48 m and 8.33 m, for source positions 1 and 2, respectively. The lowest r_D values were recorded when both the water sound and air-conditioning system were in operation, with r_D values as low as 1.70 m for source position 1 and 2.05 m of source position 2. The water sound resulted in the second lowest r_D values. The values of r_D were 3.74 m for source position 1 and 4.26 m for source position 2.

The energy averaged A-weighted background noise in the space was 36.51 dB for the no masking condition and rose to 43.89 dB and 44.29 dB for the water sound, and air-conditioning noise, respectively. When both the water sound and air-conditioning noise

were combined, the resulting energy averaged A-weighted background noise level was 46.58 dB. The background noise levels created by the masking sounds were all within the acceptable range of masking sound level (see Chapter 4). It is worth mentioning that the average distraction distance (i.e., taking values from both sound source positions) caused by the water sound was almost half that of the air-conditioning noise, despite having a similar sound pressure level. On average, the distraction area caused by the water sound was 50.26 m², while the air-conditioning noise resulted in a distraction distance as large as 196.31 m². This means that fewer people would be distracted when irrelevant speech in this space was masked by the water sound in comparison to the air-conditioning noise.

Table 8.4 Single-number quantities measured in Office 2 from two source positions.

Source position 1 (7 microphone positions)				
Masking condition	NM	AI	AW	WA
STI at the nearest workstation	0.77	0.75	0.67	0.70
Distraction distance, r_D , in m	13.24	7.48	1.70	3.74
Privacy distance, r_P , in m (extrapolated)	35.14	24.24	16.70	20.22
Spatial decay rate of A-weighted SPL of speech, $D_{2,S}$, in dB	-3.59	-3.59	-3.59	-3.59
A-weighted SPL of speech at 4 metres, $L_{p,A,S,4\text{ m}}$, in dB	53.70	53.70	53.70	53.70
Average A-weighted background noise level, $L_{p,A,B}$, in dB	36.51	44.29	46.58	43.86
Source position 2 (6 microphone positions)				
Masking condition	NM	AI	AW	WA
STI at the nearest workstation	0.72	0.68	0.60	0.62
Distraction distance, r_D , in m	14.31	8.33	2.05	4.26
Privacy distance, r_P , in m (extrapolated)	36.86	26.08	18.01	21.50
Spatial decay rate of A-weighted SPL of speech, $D_{2,S}$, in dB	-3.42	-3.42	-3.42	-3.42
A-weighted SPL of speech at 4 metres, $L_{p,A,S,4\text{ m}}$, in dB	53.90	53.90	53.90	53.90
Average A-weighted background noise level, $L_{p,A,B}$, in dB	36.51	44.29	46.58	43.86

NM: No masking. AI: Air-condition noise. AW: Air-condition noise and water sound. WA: Water sound.

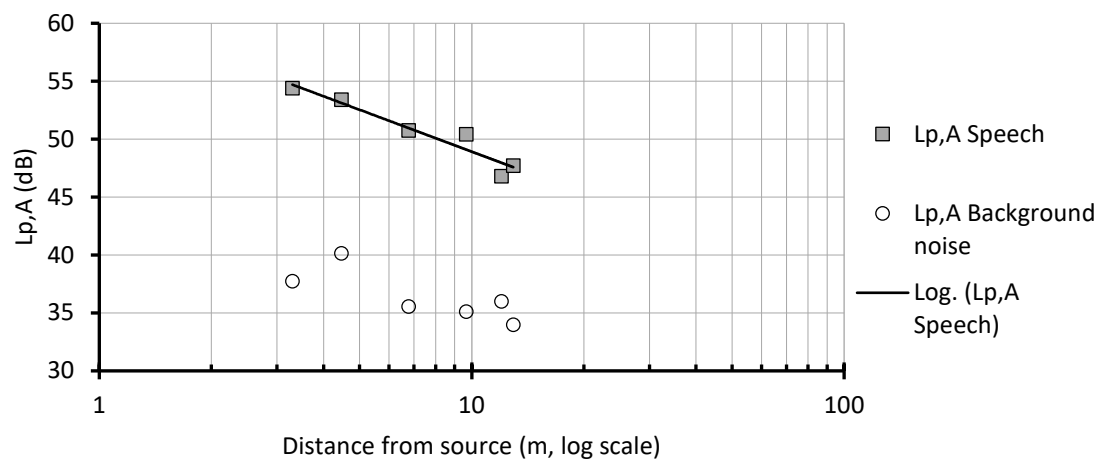


Figure 8.10 Spatial decay rate of the A-weighted sound pressure level of speech in Office 2 from source position 1. Data markers: measured data; trend line: logarithmic regression line.

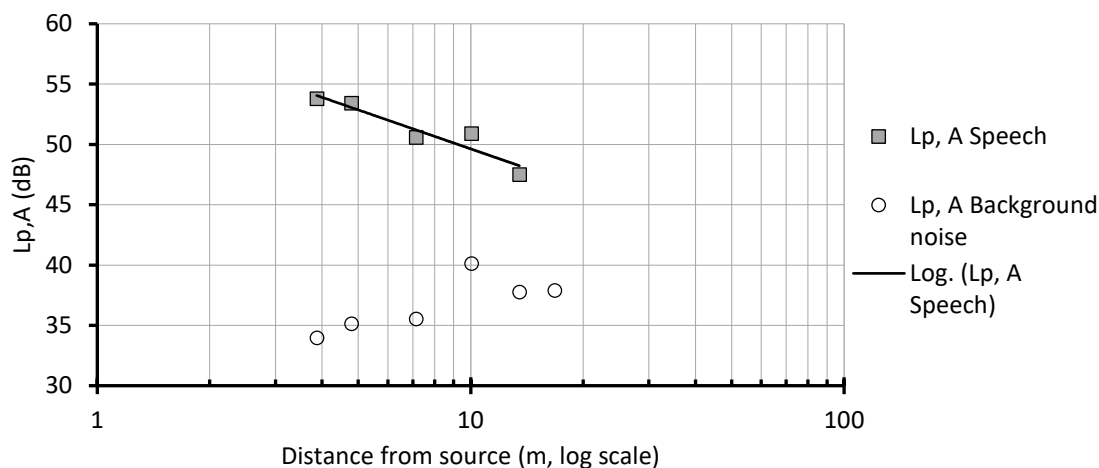


Figure 8.11 Spatial decay rate of the A-weighted sound pressure level of speech in Office 2 from source position 2. Data markers: measured data; trend line: logarithmic regression line.

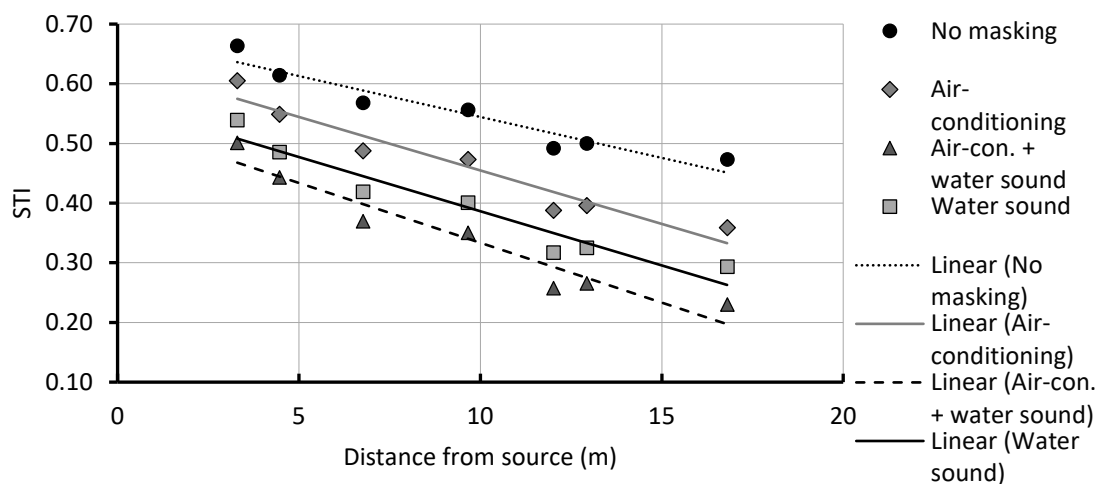


Figure 8.12 Rate of drop in STI in Office 2 measured for source position 2, in four masking conditions: no masking, air-conditioning noise, water sound, air-conditioning + water sounds. Data markers: measured data; trend lines: linear regression line.

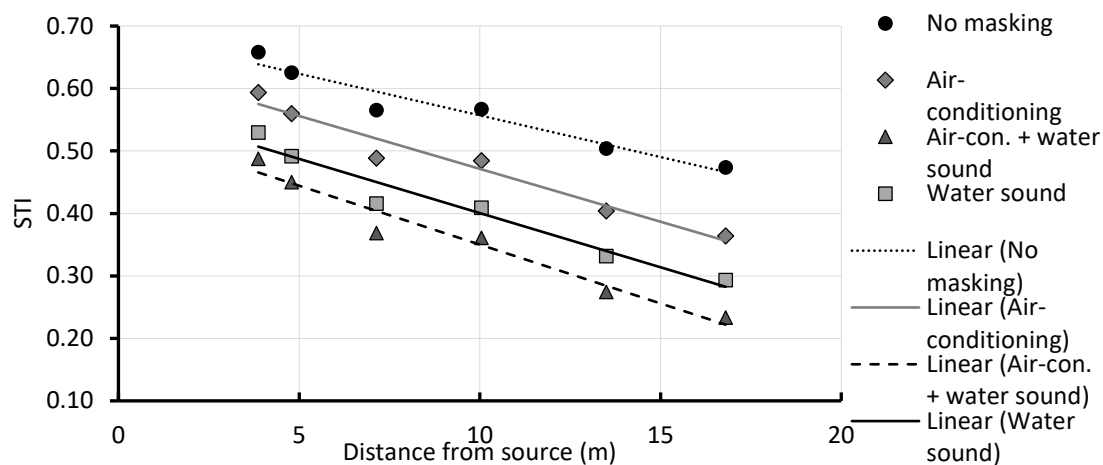


Figure 8.13 Rate of drop in STI in Office 2 measured for source position 2, in four masking conditions: no masking, air-conditioning noise, water sound, air-conditioning + water sounds. Data markers: measured data; trend lines: linear regression line.

8.8 Discussion

The addition of the water sound in both open-plan offices resulted in a drop in the value of r_D . The drop was between 8.64 m and 9.14 m for Office 1 and between 9.50 m and 10.05 m for Office 2. The air-conditioning noise in Office 2 reduced the r_D by only 5.76 m to 5.98 m, despite having a similar sound pressure level to the water sound. The greater drop caused by the water sound is likely to have happened due to the water sound containing more high-frequency sounds which might have caused a higher level of masking, despite having an overall similar sound pressure level as the air-conditioning noise.

Haapakangas *et al.* (2017) suggested that an increase of one meter in the distraction distance is associated with 9% to 14% increase in the annoyance level, and these ratios increase as the power function of the distance (see Section 2.13). Based on this suggestion, Figure 8.14 has been created in this study which shows the expected increase in the annoyance level as a function of the change in the distraction distance. According to Figure 8.14, the addition of the water sound would likely result in a reduction in the annoyance level between 190% to 550% for Office 1 and between 175% to 480% for Office 2, depending on the source position, and whether the 9% curve or the 14% curve is assumed.

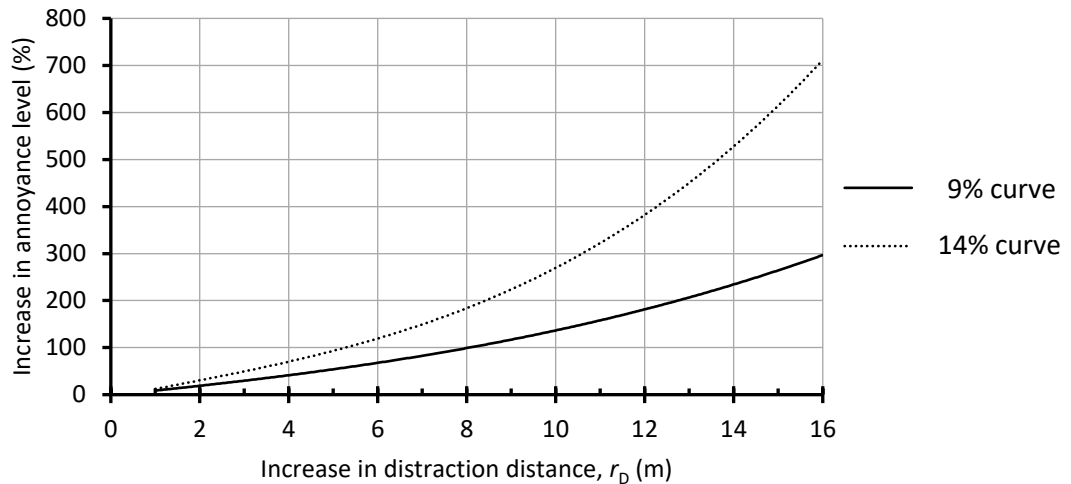


Figure 8.14 Increase in the annoyance level as a function of the increase in the distraction distance. Trendlines created based on findings reported in Haapakangas *et al.* (2017).

The highest drop in r_D was recorded when both the water sound and air-conditioning noise were present in Office 2. The octave-band spectra of the water sound and the air-

conditioning measured in Office 2 are shown in Figure 8.15. The figure also shows the masking spectrum recommended by Veitch *et al.* (2002) (i.e., a pink noise whose sound pressure level decreases by 5 dB per octave band). When both the water sound and the air-conditioning noise were combined, the resulting spectrum became very similar to that of the recommended masking spectrum. These results suggest that, in this space, the water sound resulted in a smaller r_D in comparison to the air conditioning noise, which is the masking sound often suggested in previous research (e.g., Veitch *et al.* 2002). However, to achieve a greater reduction in r_D , both the air-conditioning noise and the water sound had to be present in the space. The use of the water sound as a speech masking sound has been subjectively validated in previous stages of the current study, however, the use of the water sound combined with air-conditioning noise will need to be subjectively examined in future research, in order to achieve more conclusive findings regarding whether a combined masking sound is more preferred than the water sound alone. It might be possible to install an artificial masking system alongside the water sound to get a greater benefit from the noise masking, and this is a topic for future research.

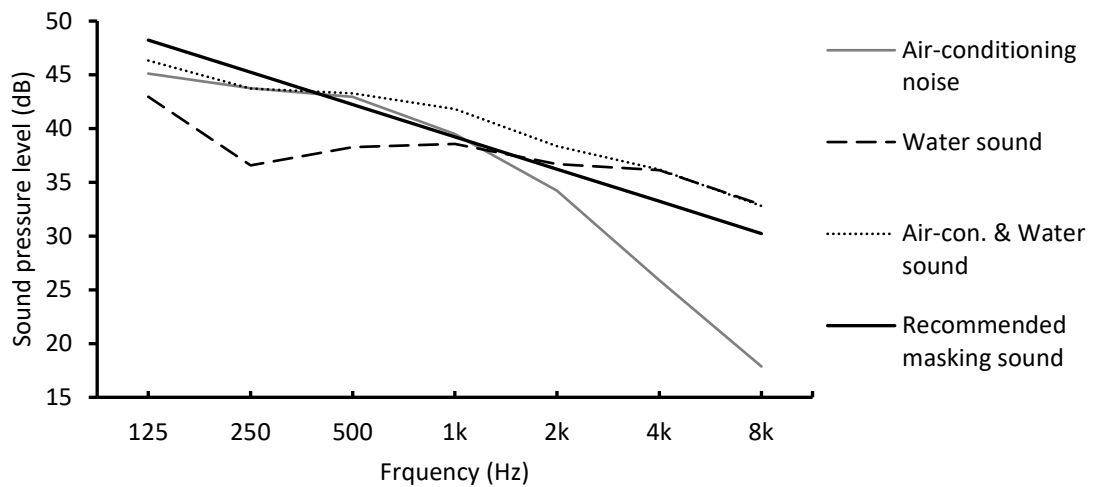


Figure 8.15 Octave-band spectra of air-conditioning noise (44.3 dBA), water sound (43.9 dBA) and air-conditioning noise and water sound combined (46.6 dBA), measured in Office 2. The recommended masking spectrum (45 dBA) is also shown for comparison purposes (Veitch *et al.*, 2002).

It appears that the two spaces did not equally benefit from the reduction in their r_D values. Despite the significant drop in r_D in both offices, the addition of the water sound did not seem to have made a significant improvement in Office 1. The distraction distances were

converted to distraction areas² for each workstation and plotted on the plan of both spaces. This plot allows for examining the extent of overlapping of the distraction areas of each workstation. Ideally, no overlapping is preferred to avoid people from different workstations distracting each other. The overlap of the distraction areas from workstations in Office 1 is shown in Figure 8.16, before and after the addition of the water sound. Darker shades represent higher levels of overlapping, and hence more distraction. Before the addition of the water sound, there was a very high level of overlapping of the distraction areas, so that almost all distraction areas overlapped each other. When the water sound was added, improvements could be noticed, as lighter shades started to appear, meaning less overlapping of the distraction areas. However, the lighter areas were located in those parts of the space where there was almost no workstation, which makes the benefit of reducing r_D in the space very limited. The distraction areas around the central part of the space, where most workstations were, remained heavily overlapping each other.

On the other hand, in Office 2, the benefit of reducing the r_D was much larger. The overlap of the distraction areas from workstations in Office 2 is shown in Figure 8.17, before and after the addition of the water sound. Generally, after introducing the water sound to the space, the overlapping of the distraction areas became much smaller, as shown by the lighter shades, and the darker shades mainly appeared in the middle of the spaces where there were fewer workstations. This could have happened due to the layout and the room acoustics of the space, as well as using partition screens between workstations.

These results suggest that the benefit of reducing the distraction distance associated with adding a masking sound to a space is likely to be affected by the layout and the acoustic of the space. The same amount of reduction in r_D might not have similar effects in different spaces. This was tested for two spaces in the current study, but more spaces will need to be tested to achieve more conclusive findings.

² Distraction area = $r_D^2 \times \pi$

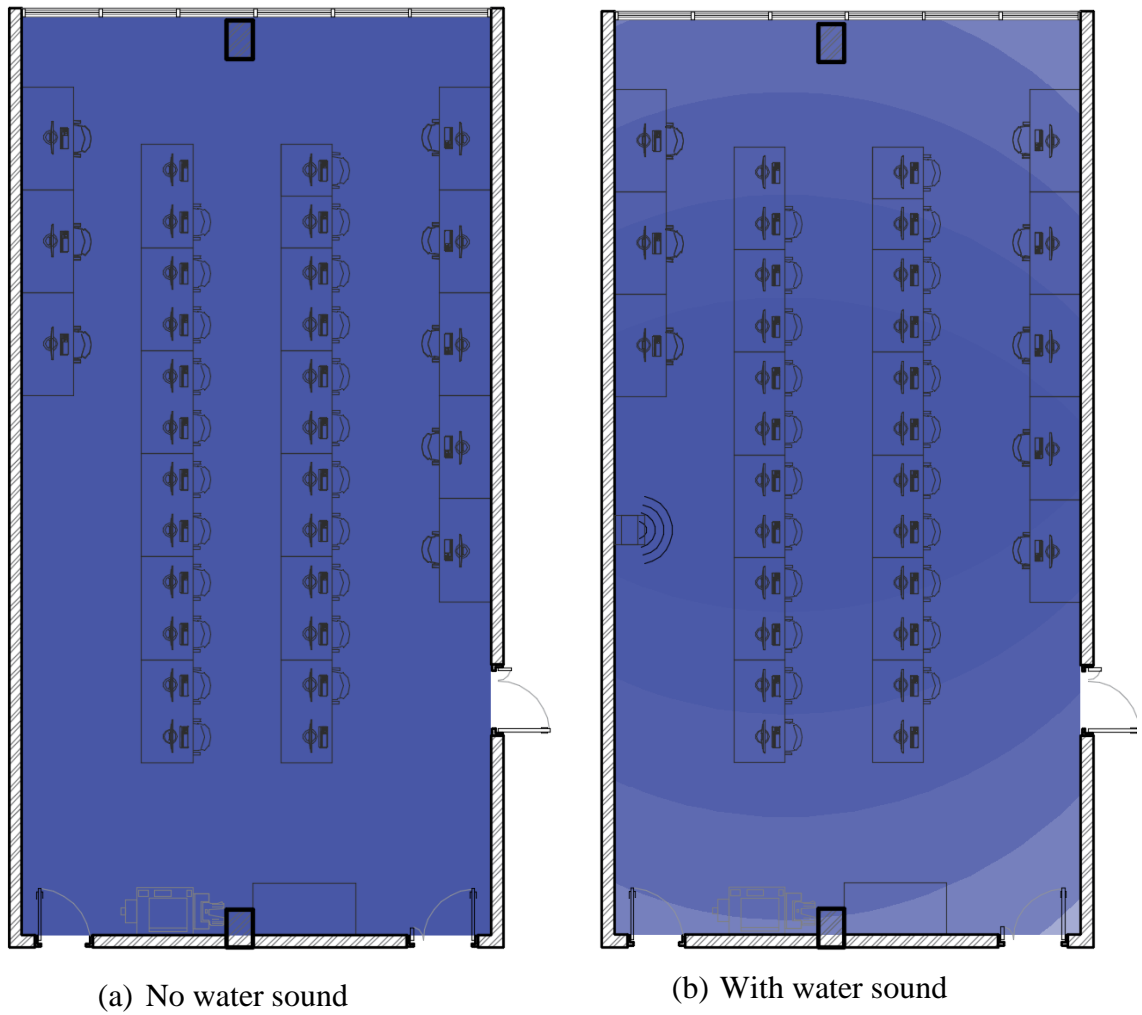


Figure 8.16 Overlapping of distraction areas of workstations in Office 1. Darker shades represent higher levels of overlapping.

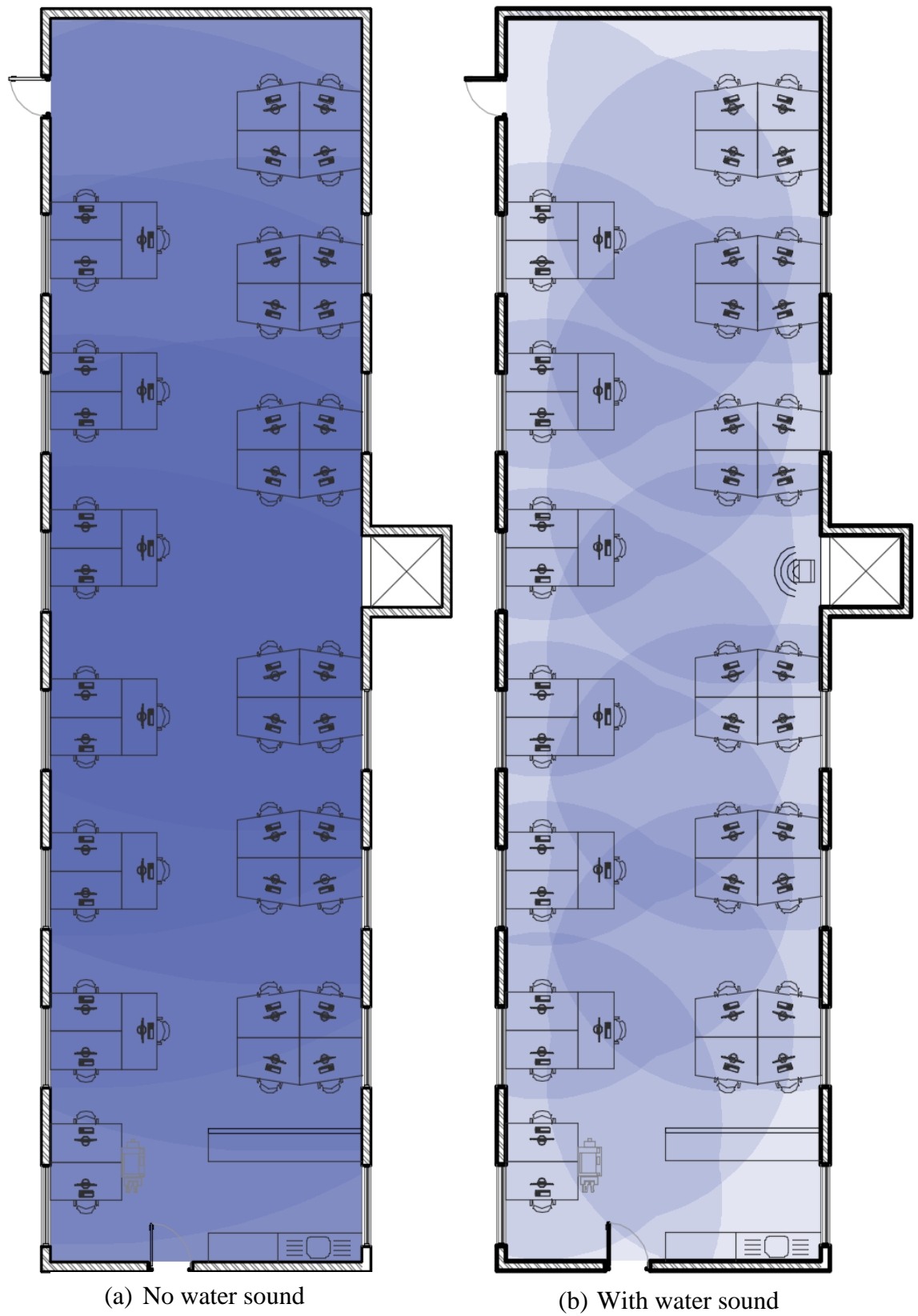


Figure 8.17 Overlapping of distraction areas of workstations in Office 2. Darker shades represent higher levels of overlapping.

8.9 Conclusions

Single-number quantities recommended by BS EN ISO 3382-3 (2012) are objective measures representative of room acoustic parameters in open-plan offices. These quantities were measured in two open-plan offices (Office 1 and Office 2) at the Edinburgh campus of Heriot-Watt University. The distraction distance, r_D and the privacy distance, r_P , which are two of the single-number quantities, were measured in the spaces before and after adding a water sound, and the drops in magnitude of r_D and r_P were calculated. The results showed that both spaces had poor office acoustics with values of the spatial decay rate of A-weighted SPL of speech, $D_{2,s}$, of around -3.5 dB, and the A-weighted SPL of speech at 4 metres, $L_{p,A,S,4m}$, of around 53 dB.

The addition of the water sound reduced the value of r_D by 8.64 m to 9.14 m in Office 1, and by 9.50 m to 10.05 m in Office 2. The results also showed that the two spaces might not have benefited similarly from the reduction in their respective r_D . The distraction areas of workstations in Office 1 remained mainly overlapping each other even after adding the water sound, while in Office 2, this overlap was significantly reduced by the water sound. This suggested that the layout and the room acoustic characteristic of the spaces might have affected the extent to which they benefited from the reduction in their r_D .

The water sound in Office 2 resulted in a greater reduction in r_D in comparison to the reduction achieved by using air-conditioning noise, despite having a similar SPL. This suggests that, objectively, the water sound was a better speech masker than the air-conditioning noise, which is often recommended as a speech masking sound in open-plan offices. This is likely to have happened due to the water sound containing higher levels of high-frequency sounds, which could have resulted in smaller STI values. The highest reduction in r_D in Office 2 was achieved when both the water sound and air-conditioning noise were played together. Therefore, it might be possible to combine both sounds to obtain a greater reduction in r_D , and possibly higher subjective satisfaction levels. However, this will need to be subjectively tested, which makes it a rich topic for future research. In the next chapter, general conclusions, impact of the research and limitations of the current study are presented, followed by recommendations for future work in the field.

CHAPTER 9: CONCLUSIONS

9.1 Introduction

In this chapter, the key findings of the current study are presented. A summary of conclusions is provided for each chapter, followed by a discussion on the impact of the research. Where necessary comparisons are drawn between results achieved in the current study and previous research, which support or contradict the findings of the current research. Limitations of the study and suggestions for future work are presented at the end of the chapter.

9.2 Findings

The main findings of the research are listed below.

- Carefully designed water features can be effective speech maskers in open-plan offices.
- The preferred sound pressure level of a water sound when used as a speech masking sound is 45 dBA.
- The preferred sound masking level of water sounds is independent from the type of the water sound and the intelligibility level of background speech.
- The preferred water sound, in an audio-only context, was a 37-jet fountain.
- The preferred water feature, in an audio-visual context, was a 4-step cascade.
- Visual stimuli have a positive effect on people's perception of their work environment.
- There are 1.1 to 2.5 more chances to make positive changes in people's perception of their work environment, when a water sound is accompanied by an appropriate visual stimulus.
- Subjective workload in open-plan offices reduces when irrelevant speech is masked by an appropriate water sound.
- Subjective speech privacy increases in open-plan offices when irrelevant speech is masked by an appropriate water sound.
- Cognitive performance in a serial recall task (short-term memory) increased by 5% when background irrelevant speech was masked by a water sound.

- Both reaction time and response accuracy should be taken into account when the cognitive performance is measured.
- The gender of the participant can have a significant impact on the cognitive performance of a serial recall task.
- A carefully designed water feature can significantly enhance the quality and pleasantness of the sound environment as well as some aspects of the physical work environment in open-plan offices.
- Water features can be a cost-effective substitute for artificial masking systems.
- A water sound can reduce the distraction distance in some open-plan offices by up to 10 m.
- Different spaces benefit differently from the reduction in the distraction distance associated with adding a water feature to them.

9.3 Conclusions

This research examined the effectiveness of using water features in open-plan offices to mask irrelevant speech and create a pleasant work environment that promote speech privacy and cognitive performance. Given the wide area that the research was covering, five experiments were carried out, each dedicated to meet a certain objective of the study. The objectives of the study were as follows:

Objective 1 - Identifying the preferred configurations of water features to be used in open-plan offices to promote speech privacy. This is achieved through the following sub-objectives:

- a) Identifying the preferred sound pressure level (SPL) of water sounds in open-plan offices as a function of speech intelligibility (STI).
- b) Identifying the preferred water features and their perception through audio-only and audio-visual tests.

Objective 2 - Examining the extent to which water sounds improve cognitive task performance as well as subjective workload and satisfaction.

Objective 3 - Examining the longer-term effects of water feature on people's perceptions.

Objective 4 - Evaluating the improvement in objective speech privacy associated with installing a water feature in open-plan offices through the measurement of distraction and privacy distances.

The five experiments, despite some limitations, enabled the current study to meet its aim and objectives as thoroughly as possible, within the allocated time and budget.

9.3.1 Conclusions of the chapters

Chapter 2 presented a review of the literature that provided necessary knowledge to carry out the current study. The literature identified irrelevant speech as a major source of annoyance in open-plan offices, and highlighted the effectiveness of using speech masking strategies in mitigating the adverse effects of irrelevant speech. Studies were reviewed where water sounds were effectively used as road traffic masking sounds. In addition, previous research highlighted water sounds as potential speech masking sounds that might outperform the currently used artificial masking sounds in offices.

Chapter 3 provided an overview of the methodologies adopted across the five experiments carried out in this research. The water sounds used in the current study were described and their acoustic and psychoacoustic properties were presented. This was followed by a brief description of the speech recording which was used as a source of irrelevant speech throughout this research. The five experiments and a general overview of their methodologies were then discussed. Lastly, the statistical models used for data analysis were provided.

Chapter 4 was dedicated to Experiment 1 “the sound level preference test” which met Objective 1.a. This experiment identified the preferred sound pressure level of water sounds when used as speech masking sounds. The results revealed the preferred sound pressure level to be 45 dBA, followed by 48 dBA and 42 dBA, respectively. In addition, the preferred masking level was found to be independent from the speech intelligibility level of irrelevant speech, as well as the type of water sound used to mask irrelevant speech.

Chapter 5 was dedicated to Experiment 2 “audio-only and audio-visual preferences and perception”, which met Objective 1.b. This chapter looked at identifying the audio-only and audio-visual preferences of people towards water sounds/features when used to mask irrelevant speech in open-plan offices. Six water features were used. The water features were a 4-step cascade (CA), a dome fountain (DF), a foam fountain (FF), a 37-jet fountain (FTW), a large jet (LJT) and a narrow jet (NJT).

The audio-only preference test revealed FTW as the most preferred water sound and NJT as the least preferred water sound. In the audio-visual preference test, CA was the most preferred water feature whilst NJT was the least preferred water feature. The feature that

most benefited from its visual stimulus was FF, whilst NJT benefited the least from its visual animation, followed by FTW. The addition of the visual stimuli increased the likelihood of making a positive change in people's perception. The odds ratios showed that there was 1.1 to 2.5 times more chance to make positive changes in perception when the audio materials were accompanied by visual animations.

Chapter 6 was dedicated to Experiment 3 "cognitive performance and subjective satisfaction", which met Objective 2. This experiment examined the effect of masking irrelevant speech on cognitive performance and subjective satisfaction, when a water sound is used as a speech masking sound. Cognitive performance of participants in four tasks, as well as their subjective satisfaction were measured under two background noise conditions; a speech-only condition and a masked speech condition. The four cognitive tasks were a serial recall task, an information matching task, a one-back task, and a reading comprehension task. For each task, participants' response accuracy (RA) and reaction time (RT) were measured.

Cognitive performance in the serial recall task was significantly affected by the background noise condition. This result was achieved when both RA and RT were simultaneously taken into account. Participants' response accuracy increased, and their reaction time dropped when irrelevant speech was masked by the water sound. No statistically significant effect of the background noise condition was detected on the cognitive performance of the remaining three tasks.

In the serial recall task and the one-back task, the effect of gender was significant. Male participants showed a significantly higher performance level, in comparison to their female counterparts. In addition, males maintained a steady performance across the two background noise conditions, while females tended to have a lower performance level in the speech-only condition, and a significantly higher performance in the masked speech condition.

The nationality of participants (native vs. non-native English speakers) on the other hand, had a significant main effect on performance in the one-back task and the reading comprehension task. Native English speakers had a shorter reaction time than non-native English speakers in the one-back task. Similarly, native English speakers' performance scores were higher than that of non-native speakers in the reading comprehension task.

Subjectively, overall satisfaction with the sound environment significantly increased after masking irrelevant speech with the water sound. Furthermore, subjective speech intelligibility and workload were significantly lower in the masked speech condition.

Chapter 7 was dedicated to Experiment 4 “longer-term exposure to a water sound” which met Objective 3. The experiment examined the longer-term effects of installing a water feature in an open-plan office. The results showed that participants were initially moderately satisfied with the comfort/functionality, and control/privacy aspects of the work environment. Including the water feature in the space significantly improved the comfort/functionality aspect of the physical work environment.

The perceived frequency of hearing intelligible speech significantly reduced after the addition of the water feature, while an insignificant increase in the perceived frequency of hearing unintelligible speech was observed. The water sound significantly improved the way participants perceived their sound environment. The pleasantness level of the sound environment and the possibility of concentrating on tasks both increased significantly after installing the water feature. Furthermore, the perceived distraction level in the space was significantly reduced by the water sound.

Chapter 8 was dedicated to Experiment 5 “objective measures” which met Objective 4. This experiment focused on examining the rate of reduction in the distraction distance, r_D , associated with installing a water feature in an open-plan office. The distraction distance was measured in two open-plan offices, Office 1 and Office 2, before and after adding a water sound to the background noise. The reduction in r_D after adding the water sound was between 8.64 m and 9.14 m in Office 1, and between 9.50 m and 10.05 m in Office 2. Despite the similarity between the rates of drop in r_D in the two spaces, Office 2 benefited more from the reduction in the distraction distance, mainly because of the layout of the space and having partitions which reduced the degree of overlapping the distraction areas between the workstations. Hence, the same rate of drop in the distraction distance might not equally benefit different spaces.

9.4 Impact of the research

Various experiments were carried out in view of making the outcomes of this research as comprehensive as possible. The findings have been validated both subjectively and objectively, and have been backed up by advanced statistical models alongside complex modelling and simulation tools. Therefore, the findings of the current study are highly

credible, and indeed, impactful. Overall the results support using water features in open-plan offices to reduce the constant noise issues and create a pleasant work space where productivity is promoted. Carefully designed water features that comply with findings and suggestions made in this study can substitute the conventional and costly masking systems, with the added benefit of being affordable, more aesthetically appealing and creating a better congruence between the acoustic space and the visual space. The findings can also be used by architects, interior designers and acoustic engineers to design work places where water is an integrated part of the design process. The research also contributes to the literature by providing much needed recommendations concerning the selection criteria of water sounds when used as speech masking sounds. This will allow future research to avoid using existing recommendations which are valid only for artificially generated sounds.

9.5 Limitations and future work

Despite the promising results, the findings of this research should not be taken out of their context. A water feature in an open-plan office is effectively a masking system that can, to a limited extent, overcome noise issues, especially lack of speech privacy. Adding a water feature to an open-plan office is like adding “perfume” to the environment. This perfume can make the soundscape more pleasant and aesthetically more appealing, however, not all noise problems can be treated by just adding some perfume; “no amount of perfumery can cover up a stinking job” (Schafer, 1994, p. 224). A water feature is not a substitute for the well-known room acoustic treatments such as having the right amount of absorption, high partition screens and appropriate height-length-depth ratio. As Virjonen *et al.* (2009) suggested, speech masking and other room acoustic treatments can collectively create a better work environment.

It also cannot be said that the water sounds used in this study outperformed the recommended artificial masking sounds. Having a direct comparison between the highly preferred water sounds and the masking sound recommended by Veitch *et al.* (2002) would be an interesting topic for future research. What is clear though, is that due to not having any visual representatives, the artificially masking sounds would clearly underperform the water sounds in terms of adding aesthetic value to the work environment as well as providing the congruence between the acoustic and visual spaces that the water features would.

In experiment 2, only two STI levels, STI 0.50 and STI 0.78, were included, which were shown to have no effect on the preferred masking level of the water sounds. The lower STI level of 0.50 might have still been considered high enough, which might justify the similarity in preference scores between the two STI conditions. Hongisto (2005) suggested that at an STI of 0.50 nearly maximum performance decrement has already occurred, while Ebissou *et al.* (2015) suggested an STI of 0.45, and Jahncke *et al.* (2013) an even lower value of STI of 0.34, where the maximum performance decrement occurs. It seems that at STI values lower than 0.50, important changes happen that could eventually have an effect on the preferred masking level. Future studies, therefore, should examine lower STI values to understand how the intelligibility level of irrelevant speech affects the preferred masking level of water sounds. Although it would be argued that speech masking systems are not needed at lower STI levels.

More attention should be given to the visual material used to represent water features. Moving water cannot be represented via still images, and hence, still images should be avoided in future research. A great amount of work has been put in the current study to create animations of water features that looked as realistic as possible. However, this could further be improved. The animations were rendered from a single perspective, yet current technology allows them to be modified so that they could be used in a virtual reality setting to create an immersive virtual environment, where the perspective changes according to the movement of the head and eyes of the observer.

It is still not clear what makes a certain water sound to be preferred. Evocative and semantic characteristics of water sounds can help, to a limited extent, in understanding people's preferences towards water sounds. In the current study, attempts were made to relate the preference ratings of water sounds to more objective measures such as the sound interference level (SIL), the low-high frequency A-weighted level difference ($L_{AeqLow} - L_{AeqHigh}$), temporal variations ($L_{10} - L_{90}$), low frequency contents ($L_{Ceq} - L_{Aeq}$), as well as psychoacoustic measures such as the sharpness, roughness and pitch strength. These measures, except roughness, and to a lesser extent, pitch strength, were unreliable in predicting the preference ratings of water sounds, mainly because of having clusters of people with opposing preferences towards water sounds. Brain imaging technologies might further advance the understanding of how different water sounds and visuals stimulate the brain which would ultimately dictate the preference ratings.

The lack of having a baseline condition in Experiment 3 “cognitive performance and subjective satisfaction” (i.e., performance in quiet) is another limitation of the current study, which posed some statistical restrictions. Having a baseline condition would have helped further advance the statistical analysis by including peoples’ response and performance in this condition as a covariate, which would have explained some variations that were treated as error in the statistical models used in the current study. This would have resulted in more statistically significant results being achieved, as the statistical models (e.g., ANOVA and MANOVA) would have been more powerful in detecting smaller effects and rendering them as being statistically significant. The baseline condition was excluded in the current study to keep the test under 50 minutes, in order to avoid participants being too fatigued. It is also well documented that people perform better in quiet. Future research could use one task, for instance, a serial recall task, and examine more background noise conditions without extending the length of the study.

It was also found that in the previously published studies, there was a tendency of over reliance on the *p*-value, and possibly misinterpreting its value. The *p*-value merely shows whether an effect has happened due to experimental manipulations or simply by chance. Its value is strongly dependent on the sample size being tested. In small samples, which is often the case in relevant soundscape studies, even large effects might not cause a statistically significant result, while in very large samples, even very small effects, and thus practically meaningless effects, can be rendered as being statistically significant. Some of the experimental manipulations across this research tended to show a small effect size, especially in the cognitive task performance experiment, and thus were statistically insignificant. However, these small effects might be practically meaningful. For instance, a 1% increase in performance can result in substantial economic savings. To overcome this problem in future research, large numbers of participants should be used, especially if an independent measure design was adopted. This would make the statistical models more powerful in detecting smaller changes and make the small effect sizes statistically significant.

Taking both reaction time and response accuracy is of paramount importance to future studies. These were shown to interact differently with the background noise conditions as well as with each other depending on the cognitive task being tested. The effect of the gender of participants should also be given particular attention in future research. In the current study, females seemed to benefit more from the masking system, but no conclusive explanation could be provided regarding why this has happened.

In experiment 4 “longer-term exposure to a water sound”, the small sample size ($N = 14$) is a limitation that makes it difficult to generalise the findings beyond the sample. In addition, the sample mainly consisted of females ($N=12$), which made comparisons between males and females not possible. The experiment was also based on a relatively small and quiet open-plan office, whose people were already satisfied before installing the water feature. It is not clear if similar findings would be achieved in more crowded and larger offices with less satisfied people. Future studies should include larger open-plan offices, with a higher number of participants. Installing multiple water features in large spaces might then be required.

An active water feature system is also a great topic for future research, which adds technology to the water sounds. Given that the acoustic properties of water features, such as loudness and sound level, can be predicted (Galbrun and Ali, 2013), active water features could be designed to have variable flow rates that would change according to the noise level and the intelligibility level of irrelevant speech. This would allow for having high levels of water sound when the speech intelligibility is high (single voice), and low levels of water sound when the intelligibility is already low (multiple voices).

Finally, in Experiment 5 “objective measures”, only two spaces were examined, both of which had poor office acoustics. It is not clear how water sounds might behave, in terms of reducing the distraction distance, in acoustically well designed open-plan offices. In addition, as Virjonen *et al.* (2009) suggested, the distraction distance can vary dramatically between open-plan offices with different designs and layouts. Hence, examining more offices in future studies would be a step forward in further validating the results achieved in the current research.

REFERENCES

- Abbott, L. C. (2015) *The Influence of Natural Sounds on Attention Restoration*. MSc dissertation, The Pennsylvania State University: Department of Recreation, Park and Tourism Management, August 2015.
- Alfirevic, D. and Simonovic, A. S. (2016) 'Open-plan in housing architecture: Origin, development and design approaches for spatial integration', *Arhitektura i urbanizam*, (43), pp. 45–60.
- Ali, T. T. (2012) *Acoustical design of water features and their use for road traffic noise masking*. PhD thesis, Heriot-Watt University: School of the Built Environment, October 2012.
- Axelsson, Ö. *et al.* (2014) 'A field experiment on the impact of sounds from a jet-and-basin fountain on soundscape quality in an urban park', *Landscape and Urban Planning*, 123, pp. 49–60.
- Banbury, S. and Berry, D. . (1997) 'Habituation and dishabituation to speech and office noise', *Journal of Experimental Psychology: Applied*, 3(3), pp. 181–195.
- Beaman, C. P. and Jones, D. M. (1997) 'Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2), pp. 459–471.
- Benfield, J. A. *et al.* (2014) 'Natural sound facilitates mood recovery', *Ecopsychology*, 6(3), pp. 183–188.
- Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society. Series B*, 57(1), pp. 289–300.
- van den Berg, A. E., Koole, S. L. and van der Wulp, N. Y. (2003) 'Environmental preference and restoration: (How) are they related?', *Journal of Environmental Psychology*, 23(2), pp. 135–146.
- Berger, C. C. and Ehrsson, H. H. (2013) 'Mental imagery changes multisensory perception', *Current Biology*, 23(14), pp. 1367–1372.
- Bistafa, S. R. and Bradley, J. S. (2000) 'Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics', *The Journal of the Acoustical Society of America*, 107(2), pp. 861–875.
- Blanchflower, D. G. and Oswald, A. (2017) *Do humans suffer a psychological low in midlife? Two approaches (with and without controls) in seven data sets*. Report Series Reference No 23724. Cambridge, Massachusetts.
- Bodin Danielsson, C. *et al.* (2014) 'Office design's impact on sick leave rates', *Ergonomics*, pp. 139–147.
- Bodin Danielsson, C. and Bodin, L. (2009) 'Difference in satisfaction with office environment among employees in different office types', *Journal of Architectural and Planning Research*, pp. 241–257.

- Bosco, A., Longoni, A. M. and Vecchi, T. (2004) 'Gender effects in spatial orientation: Cognitive profiles and mental strategies', *Applied Cognitive Psychology*, 18(5), pp. 519–532.
- Bradley, J. S. (1986a) 'Auditorium acoustics measures from pistol shots', *The Journal of the Acoustical Society of America*, 80(1), pp. 199–205.
- Bradley, J. S. (1986b) 'Speech intelligibility studies in classrooms', *The Journal of the Acoustical Society of America*, 80(3), pp. 846–854.
- Bradley, J. S. (2003) 'The Acoustical design of conventional open-plan offices', *Canadian Acoustics*, 31(2), pp. 23–31.
- Bradley, J. S., Reich, R. Da and Norcross, S. G. (1999a) 'A just noticeable difference in C50 for speech', *Applied Acoustics*, 58(2), pp. 99–108.
- Bradley, J. S., Reich, R. Da and Norcross, S. G. (1999b) 'On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility', *The Journal of the Acoustical Society of America*, 106(4), pp. 1820–1828.
- Brocolini, L., Parizet, E. and Chevret, P. (2016) 'Effect of masking noise on cognitive performance and annoyance in open-plan offices', *Applied Acoustics*, 114, pp. 44–55.
- Brown, L. and Rutherford, S. (1994) 'Using the sound of water in the city', *Landscape Australia*, 2, pp. 103–107.
- BS EN ISO 3382-1 (2009) *Acoustics-Measurement of room acoustic parameters-Part 1: Performance spaces*. Brussels: European Committee for Standardization.
- BS EN ISO 3382-2 (2008) *Incorporating corrigendum Acoustics-Measurement of room acoustic parameters-Part 2: Reverberation time in ordinary rooms*. Brussels: BSI Standards Limited.
- BS EN ISO 3382-3 (2012) *Measurement of room acoustic parameters-Part 3: Open-plan offices*. BSI Standards Limited.
- BS EN ISO 3741 (2010) *Acoustics - Determination of sound power levels and sound energy levels of noise sources using sound pressure - Precision methods for reverberation test rooms*. Brussels: BSI Standards Limited.
- BS EN ISO 9921 (2003) *Ergonomics - Assessment of speech communication*. Brussels: BSI Standards Limited.
- BS ISO 12913 (2014) *Acoustics-Soundscape Part 1: Definition and conceptual framework*. Geneva: BSI Standards Limited.
- Burmil, S., Daniel, T. C. and Hetherington, J. D. (1999) 'Human values and perceptions of water in arid landscapes', *Landscape and Urban Planning*, 44(2–3), pp. 99–109.
- Caplan, P. J. *et al.* (1997) *Gender differences in human cognition*. New York: Oxford University Press.
- Carles, J., Bernáldez, F. and Lucio, J. D. (1992) 'Audio-visual interactions and soundscape preferences', *Landscape Research*, 17(2), pp. 52–56.
- De Coensel, B., Vanwetswinkel, S. and Botteldooren, D. (2011) 'Effects of natural sounds

- on the perception of road traffic noise.’, *The Journal of the Acoustical Society of America*, 129(4), pp. 148–53.
- Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*. 2nd edn. New York: Academic Press.
- De Croon, E. M. *et al.* (2005) ‘The effect of office concepts on worker health and performance: a systematic review of the literature.’, *Ergonomics*, 48(2), pp. 119–134.
- Dokmeci, P. N. and Kang, J. (2010) ‘Objective parameters for acoustic comfort in enclosed spaces’, in *International Congress on Acoustics, ICA 2010*. Sydney, Australia, pp. 1–4.
- Durlach, N. I. *et al.* (2003) ‘Note on informational masking.’, *The Journal of the Acoustical Society of America*, 113(6), pp. 2984–2987.
- Dykstra, L. (2008) *Aqua house: Architecture that integrates water technically and aesthetically*. VDM Verlag.
- Ebissou, A., Parizet, E. and Chevret, P. (2015) ‘Use of the Speech Transmission Index for the assessment of sound annoyance in open-plan offices’, *Applied Acoustics*, 88, pp. 90–95.
- Efron, B. and Tibshirani, R. J. (1993) *An introduction to the bootstrap*. New York: Chapman & Hall.
- Ellermeier, W. and Hellbruck, J. (1998) ‘Is level irrelevant in “irrelevant speech”? Effects of loudness, signal-to-noise ratio, and binaural unmasking.’, *Journal of experimental psychology. Human perception and performance*, 24(5), pp. 1406–1414.
- Faggi, A. *et al.* (2013) ‘Water as an appreciated feature in the landscape: a comparison of residents’ and visitors’ preferences in Buenos Aires’, *Journal of Cleaner Production*, 60, pp. 182–187.
- Fastl, H. (2006) ‘Psychoacoustic basis of sound quality evaluation and sound engineering’, in *The 13th International Congress on Sound and Vibration*. Vienna, Austria, July 2006, pp. 324–345.
- Fastl, H. and Zwicker, E. (2007) *Psychoacoustics - Facts and Models*. 3rd edn. New York: Springer-Verlag Berlin Heidelberg.
- Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics*. 4th edn. London: SAGE Publications Ltd.
- Fletcher, H. (1953) *Speech and Hearing in Communication*. Edited by J. B. Allen. New York: Acoustical Society of America.
- Franz, G. J. (1959) ‘Splashes as sources of noise in liquids’, *Journal of Sound and Vibration*, 31, pp. 1080–1096.
- Galbrun, L. and Ali, T. T. (2013) ‘Acoustical and perceptual assessment of water sounds and their use over road traffic noise.’, *The Journal of the Acoustical Society of America*, 133(1), pp. 227–37.
- Galbrun, L. and Calarco, F. M. A. (2014) ‘Audio-visual interaction and perceptual assessment of water features used over road traffic noise.’, *The Journal of the*

Acoustical Society of America, 136(5), pp. 2609–20.

- Galindo, M. P. and Corraliza, J. (2000) 'Environmental Aesthetics and Psychological wellbeing: Relationships between preference judgements for urban landscapes and other relevant affective responses', *Psychology in Spain*, 4(1), pp. 13–27.
- Haapakangas, A., Haka, M., *et al.* (2008) 'Effect of speech intelligibility on task performance - an experimental laboratory study', in *9th Congress of the International Commission on Biological Effects of Noise ICBEN*. Mashantucket, Connecticut, USA, July 21-25.
- Haapakangas, A., Helenius, R., *et al.* (2008) 'Perceived acoustic environment, work performance and well-being - survey results from Finnish offices', in *9th International Congress on Noise as a Public Health Problem (ICBEN)*. Mashantucket, Connecticut, USA, July 21-25.
- Haapakangas, A. *et al.* (2011) 'Effects of five speech masking sounds on performance and acoustic satisfaction. implications for open-plan offices', *Acta Acustica united with Acustica*, 97, pp. 641–655.
- Haapakangas, A. *et al.* (2014) 'Effects of unattended speech on performance and subjective distraction: The role of acoustic design in open-plan offices', *Applied Acoustics*, 86, pp. 1–16.
- Haapakangas, A. *et al.* (2017) 'Distraction distance and perceived disturbance by noise- An analysis of 21 open-plan offices', *The Journal of the Acoustical Society of America*, 141(127), pp. 127–136.
- Haka, M. *et al.* (2009) 'Performance effects and subjective disturbance of speech in acoustically different office types - A laboratory experiment', *Indoor Air*, 19(6), pp. 454–467.
- Halpern, D. F. (2000) *Sex Differences in Cognitive Abilities*. 3rd edn. New York: Taylor & Francis.
- Hannon, B. and Daneman, M. (2001) 'A new tool for measuring and understanding individual differences in the component processes of reading comprehension', *Journal of Educational Psychology*, 93(1), pp. 103–128.
- Hartig, T. *et al.* (2003) 'Tracking restoration in natural and urban field settings', *Journal of Environmental Psychology*, 23(2), pp. 109–123.
- Hedge, A. (1982) 'The Open-Plan Office: A Systematic Investigation of Employee Reactions to Their Work Environment', *Environment and Behavior*, 14(5), pp. 519–542.
- Helenius, R. and Hongisto, V. (2004) 'The effect of acoustical improvement of an open-plan office on workers', in *Proceedings of Inter-Noise 2004*. Prague, Czech Republik, August 21-25.
- Hong, J. Y. and Jeon, J. Y. (2013) 'Designing sound and visual components for enhancement of urban soundscapes.', *The Journal of the Acoustical Society of America*, 134(3), pp. 2026–36.
- Hongisto, V. (2005) 'A model predicting the effect of speech of varying intelligibility on work performance', *Indoor Air*, 15(6), pp. 458–468.

- Hongisto, V. (2008) 'Effect of sound masking on workers in an open office', in *Acoustics 08 Paris*. France, Paris, June 29- July 4, pp. 537–542.
- Hongisto, V. *et al.* (2017) 'Perception of water-based masking sounds-long-term experiment in an open-plan office', *Frontiers in Psychology*, 8(July), pp. 1–14.
- Hongisto, V., Keränen, J. and Larm, P. (2004) 'Simple model for the acoustical design of open-plan offices', *Acta Acustica united with Acustica*, 90, pp. 481–495.
- Hongisto, V., Oliva, D. and Rekola, L. (2015) 'Subjective and objective rating of spectrally different pseudorandom noises-Implications for speech masking design', *The Journal of the Acoustical Society of America*, 137(3), pp. 1344–1355.
- Hopwood, R. (2009) *Fountains and Water Features: From Ancient Springs to Modern Marvels*. London: Frances Lincoln Ltd.
- Houtgast, T. and Steeneken, H. J. M. (1973) 'The Modulation Transfer Function in Room acoustics as a Predictor of Speech Intelligibility', *Acustica*, 28, pp. 66–73.
- Houtgast, T. and Steeneken, H. J. M. (1985) 'A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria', *The Journal of the Acoustical Society of America*, 77(3), pp. 1069–1077.
- IEC 60268-16 (2011) *Sound system equipment Part 16: Objective rating of speech intelligibility by speech transmission index, Sound system equipment Part 16: Objective rating of speech intelligibility by speech transmission index*. Brussels: European Committee for Electrotechnical Standardization.
- IEC 61672-1 (2013) *Electroacoustics-Sound level meters Part 1: Specifications Measurement microphones*. Brussels: European Committee for Electrotechnical Standardization.
- ISO 15666 (2003) *Acoustics — Assessment of noise annoyance by means of social and socio-acoustic surveys*. The International Organization for Standardization.
- Jacob, K. (2001) 'Understanding Speech Intelligibility and the Fire Alarm Code', in *National Fire Protection Association*. California, USA, May 14.
- Jahncke, H. *et al.* (2011a) 'Open-plan office noise: Cognitive performance and restoration', *Journal of Environmental Psychology*, 31(4), pp. 373–382.
- Jahncke, H. *et al.* (2011b) 'Open-plan office noise: Cognitive performance and restoration', *Journal of Environmental Psychology*, 31(4), pp. 373–382.
- Jahncke, H. *et al.* (2016) 'Office noise: Can headphones and masking sound attenuate distraction by background speech?', *Work*, 55(3), pp. 505–513.
- Jahncke, H., Eriksson, K. and Naula, S. (2015) 'The effects of auditive and visual settings on perceived restoration likelihood', *Noise and Health*, 17(74), pp. 1–10.
- Jahncke, H., Hongisto, V. and Virjonen, P. (2013) 'Cognitive performance during irrelevant speech: Effects of speech intelligibility and office-task characteristics', *Applied Acoustics*, 74(3), pp. 307–316.
- Jensen, K. L., Arens, E. and Zagreus, L. (2005) 'Acoustical Quality in Office Workstations , As Assessed By Occupant Surveys', in *Indoor Air: 10th international conference on indoor air quality and climate*. Beijing, China, pp. 2401–2405.

- Jeon, J. Y. *et al.* (2010) 'Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds.', *The Journal of the Acoustical Society of America*, 127(3), pp. 1357–66.
- Jeon, J. Y. *et al.* (2012) 'Acoustical characteristics of water sounds for soundscape enhancement in urban open spaces.', *The Journal of the Acoustical Society of America*, 131(3), pp. 2101–9.
- Jones, D. ., Madden, C. . and Miles, C. (1992) 'Privileged access by irrelevant speech to short term memory: the role of changing state', *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 44 (A), pp. 645–659.
- Kang, J. (2007) *Urban Sound Environment*. Oxfordshire: Taylor & Francis.
- Kaplan, R. and Kaplan, S. (1989) *The Experience of Nature: A Psychological Perspective*. Cambridge: Cambridge University Press.
- Kaplan, S. (1995) 'The restorative benefits of nature: Toward an integrative framework', *Journal of Environmental Psychology*, 15(3), pp. 169–182.
- Karmanov, D. and Hamel, R. (2008) 'Assessing the restorative potential of contemporary urban environment(s): Beyond the nature versus urban dichotomy', *Landscape and Urban Planning*, 86(2), pp. 115–125.
- Keränen, J. and Hongisto, V. (2013) 'Prediction of the spatial decay of speech in open-plan offices', *Applied Acoustics*, 74, pp. 1315–1325.
- Keus van de Poll, M. *et al.* (2015) 'Unmasking the effects of masking on performance: The potential of multiple-voice masking in the office environment', *The Journal of the Acoustical Society of America*, 138(2), pp. 807–816.
- Keus Van De Poll, M. *et al.* (2014) 'Disruption of writing by background speech: The role of speech transmission index', *Applied Acoustics*, 81, pp. 15–18.
- Kimura, D. (1999) *Sex and Cognition*. Cambridge: The MIT Press.
- Kweon, B. S. *et al.* (2008) 'Anger and stress: the role of landscape posters in an office setting', *Environment and Behaviour*, 40, pp. 355–381.
- Lamb, S. and Kwok, K. C. S. (2016) 'A longitudinal investigation of work environment stressors on the performance and wellbeing of office workers', *Applied Ergonomics*, 52, pp. 104–111.
- Landström, U. *et al.* (2002) 'Annoyance and Performance Effects of Nearby Speech', *Acta Acustica united with Acustica*, 88(4), pp. 449–553.
- Lange, E. and Schaeffer, P. V. (2001) 'A comment on the market value of a room with a view', *Landscape and Urban Planning*, 55(2), pp. 113–120.
- Laumann, K., Gärling, T. and Stormark, K. M. (2003) 'Selective attention and heart rate responses to natural and urban environments', *Journal of Environmental Psychology*, 23(2), pp. 125–134.
- Lehrman, J. B. (1980) *Earthly Paradise: Garden and Courtyard in Islam*. Berkeley and Los Angeles: University of California Press.
- Leighton, T. G. (1994) *The Acoustic Bubble*. London: Academic Press.

- Lewis, E., Sykes, D. and Lemieux, P. (2003) 'Using a web-based test to measure the impact of noise on knowledge workers' productivity', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Denver, 47(18), pp. 1972–1976.
- Liebl, A., Assfalg, A. and Schlittmeier, S. J. (2016) 'The effects of speech intelligibility and temporal-spectral variability on performance and annoyance ratings', *Applied Acoustics*, 110, pp. 170–175.
- Loewen, L. and Suedfeld, P. (1992) 'Cognitive and arousal effects of masking office noise', *Environment and Behavior*, 24(3), pp. 381–395.
- Long, M. (2014) *Architectural Acoustics*. 2nd edn. Edited by L. Moises and R. Stem. Elsevier Academic Press.
- Luttik, J. (2000) 'The value of trees, water and open space as reflected by house prices in the Netherlands', *Landscape and Urban Planning*, 48(3–4), pp. 161–167.
- Mantiuk, R., Tomaszewska, A. and Mantiuk, R. (1981) 'Comparison of four subjective methods for image quality assessment', *Computer Graphics Forum*, 31(8), pp. 1–13.
- Marsh, J. E., Hughes, R. W. and Jones, D. M. (2008) 'Auditory distraction in semantic memory: A process-based approach', *Journal of Memory and Language*, 58(3), pp. 682–700.
- Marsh, J. E. and Jones, D. M. (2010) 'Cross-modal distraction by background speech: what role for meaning?', *Noise & Health*, 12(49), pp. 210–216.
- Mäntylä, T. (2013) 'Gender Differences in Multitasking Reflect Spatial Ability', *Psychological Science*, 24(4), pp. 514–520.
- Minnaert, M. (1933) 'XVI. On musical air-bubbles and the sounds of running water', *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. Taylor & Francis, 16(104), pp. 235–248.
- Moore, B. C. J. (1995) 'Frequency analysis and masking', in Moore, B. C. J. (ed.) *Hearing*. London: Academic Press, pp. 161–205.
- Morris, N. and Jones, D. M. (1990) 'Habituation to irrelevant speech: Effects on a visual short-term memory task', *Perception and Psychophysics*, 47(3), pp. 291–297.
- Navai, M. and Veitch, J. (2003) *Acoustic satisfaction in open-plan offices: Review and recommendations*. Ottawa, Canada.
- Nilsson, M. E. *et al.* (2010) 'Auditory masking of wanted and unwanted sounds in a city park', *Noise Control Eng. Journal*, 58(5), pp. 524–531.
- Nordh, H. *et al.* (2009) 'Components of small urban parks that predict the possibility for restoration', *Urban Forestry & Urban Greening*, 8(4), pp. 225–235.
- Northcote, S. and Trevelyan, C. (1954) 'Report on the Organisation of the Permanent Civil Service', *Public Administration*, 32(1), pp. 1–16.
- Oldham, G. *et al.* (1995) 'Listen while you work? Quasi-experimental relations between personal-stereo headset use and employee work responses.', *Journal of Applied Psychology*, 80(5), pp. 547–564.

- Oldham, G. R. and Brass, D. J. (1979) 'Employee Reactions to an Open-Plan Office: A Naturally Occurring Quasi-Experiment', *Administrative Science Quarterly*, 24(2), pp. 267–284.
- Oostendorp, A. and Berlyne, D. E. (1978) 'Dimensions in the perception of architecture: identification and interpretation of dimensions of similarity', *Scandinavian Journal of Psychology*, 19(1), pp. 73–82.
- Owen, A. M. *et al.* (2005) 'N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies', *Human Brain Mapping*, 25(1), pp. 46–59.
- Pejtersen, J. *et al.* (2006) 'Indoor climate, psychosocial work environment and symptoms in open-plan offices', *Indoor Air*, 16, pp. 392–401.
- Perham, N., Marsh, J. E. and Jones, D. M. (2009) 'Syntax and serial recall: how language supports short-term memory for order', *The Quarterly Journal of Experimental Psychology*, 62(7), pp. 1285–1293.
- Peters, R. J., Smith, B. J. and Hollins, M. (2011) *Acoustics and Noise Control*. 3rd edn. London: Pearson Education.
- Pheasant, R., Horoshenkov, K. and Watts, G. (2008) 'The acoustic and visual factors influencing the construction of tranquil spaces in urban and rural environments', *The Journal of the Acoustical Society of America*, 123, pp. 1446–1457.
- Pierrette, M. *et al.* (2014) 'Noise effect on comfort in open-space offices: development of an assessment questionnaire.', *Ergonomics*, 139(January), pp. 1–11.
- Pirn, R. (1971) 'Acoustical variables in open-planning', *The Journal of the Acoustical Society of America*, 49, pp. 1339–1345.
- Rådsten-Ekman, M., Axelsson, Ö. and Nilsson, M. E. (2013) 'Effects of Sounds from Water on Perception of Acoustic Environments Dominated by Road-Traffic Noise', *Acta Acustica united with Acustica*, 99(2), pp. 218–225.
- Richardson, J. T. E. (1994) 'Gender differences in mental rotation', *Perceptual and Motor Skills*, 78, p. 435–448.
- Roche, L., Dewey, E. and Littlefair, P. (2000) 'Occupant reactions to daylight in offices', *Lighting Research and Technology*, 32, pp. 119–126.
- Sabine, W. C. (1922) *Collected Papers on Acoustics*. Cambridge: Harvard University Press.
- Salamé, P. and Baddeley, A. (1989) 'Effects of background music on phonological short-term memory', *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 41 (A)(1), pp. 107–122.
- Schafer, R. M. (1994) *The soundscape: Our sonic environment and the tuning of the world*. Vancouver: Destiny Books.
- Schlittmeier, S. and Hellbrück, J. (2009) 'Background music as noise abatement in open-plan offices: A laboratory study on performance effects and subjective preferences', *Applied Cognitive Psychology*, 23, pp. 684–697.
- Schlittmeier, S. J. *et al.* (2008) 'The impact of background speech varying in

- intelligibility: Effects on cognitive performance and perceived disturbance', *Ergonomics*, 51(5), pp. 719–736.
- Schütte, M., Marks, A., *et al.* (2007) 'The Development of the Noise Sensitivity Questionnaire', *Noise and Health*, 9(34), pp. 15–24.
- Schütte, M., Sandrock, S. and Griefahn, B. (2007) 'Factorial validity of the noise sensitivity questionnaire', *Noise and Health*, 9(37), pp. 96–100.
- Smith-Jackson, T. L. and Klein, K. W. (2009) 'Open-plan offices: Task performance and mental workload', *Journal of Environmental Psychology*, 29(2), pp. 279–289.
- Staats, H., Kieviet, A. and Hartig, T. (2003) 'Where to recover from attentional fatigue: An expectancy-value analysis of environmental preference', *Journal of Environmental Psychology*, 23(2), pp. 147–157.
- Steeneken, H. J. M. and Houtgast, T. (1980) 'A Physical Method for Measuring Speech-transmission Quality', *The Journal of the Acoustical Society of America*, 67(1), pp. 318–326.
- Stephen, S. (1988) 'Water Features: Pools, Cascades, Fountains', *The Architect's Journal*, 2.
- Sundstrom, E. *et al.* (1994) 'Office Noise, Satisfaction, and Performance', *Environment and Behavior*, 26(2), pp. 195–222.
- Sundstrom, E., Herbert, R. K. and Brown, D. W. (1982) 'Privacy and Communication in an Open-Plan Office: A Case Study', *Environment and Behavior*, 14(3), pp. 379–392.
- Tennessen, C. M. and Cimprich, B. (1995) 'Views to nature: Effects on attention', *Journal of Environmental Psychology*, 15(1), pp. 77–85.
- UCL (2017) *Speech, hearing & phonetic sciences, Chapter 6, UCL Division of Psychology and Language Sciences*. Available at: <http://www.phon.ucl.ac.uk/courses/spsci/iss/week6.php> (Accessed: 23 August 2017).
- Ulrich, R. . (1983) 'Aesthetic and affective response to natural environment', in *Behavior and Natural Environment*. New York: Plenum Press, pp. 85–125.
- Ulrich, R. S. *et al.* (1991) 'Stress recovery during exposure to natural and urban environments', *Journal of Environmental Psychology*, 11(3), pp. 201–230.
- Vassie, K. and Richardson, M. (2017) 'Effect of self-adjustable masking noise on open-plan office worker's concentration, task performance and attitudes', *Applied Acoustics*, 119, pp. 119–127.
- Veitch, J. *et al.* (2002) *Masking speech in open-plan offices with simulated ventilation noise: noise level and spectral composition effects on acoustic satisfaction*. Internal Report No. IRC-IR-846, National Research Council, Canada.
- Venetjoki, N. *et al.* (2006) 'The effect of speech and speech intelligibility on task performance.', *Ergonomics*, 49(11), pp. 1068–91.
- Virjonen, P. *et al.* (2007) 'Speech Privacy Between Neighboring Workstations in an Open Office - A Laboratory Study', *Acta Acustica united with Acustica*, 93(5), pp. 771–

- Virjonen, P., Keränen, J. and Hongisto, V. (2009) 'Determination of Acoustical Conditions in Open-Plan Offices: Proposal for New Measurement Method and Target Values', *Acta Acustica united with Acustica*, 95, pp. 279–290.
- Warnock, A. C. C. (1973) 'Acoustical privacy in the landscaped office', *The Journal of the Acoustical Society of America*, 53, pp. 1535–1542.
- Watson, C. S. (2005) 'Some comments on informational masking', *Acta Acustica united with Acustica*, 91(3), pp. 502–512.
- Watts, G. R. *et al.* (2009) 'Measurement and subjective assessment of water generated sounds', *Acta Acustica united with Acustica*, 95, pp. 1032–1039.
- Wells, N. . (2000) 'At home with nature: effects of "greenness" on children's cognitive functioning', *Environment and Behavior*, 32, pp. 775–795.
- West, M. and Parkin, P. H. (1975) 'Effect of furniture and boundary conditions on the sound attenuation in a landscaped office. Part 1', *Applied Acoustics*, 8, pp. 43–66.
- West, M. and Parkin, P. H. (1978) 'Effect of furniture and boundary conditions on the sound attenuation in a landscaped office. Part 2', *Applied Acoustics*, 11(3), pp. 171–2018.
- White, M. *et al.* (2010) 'Blue space: The importance of water for preference, affect, and restorativeness ratings of natural and built scenes', *Journal of Environmental Psychology*, 30(4), pp. 482–493.
- WHO (2000) *Noise and health. - In: Local authorities, health and environment briefing pamphlet series, 35.* Edited by B. Bonnefoy and M. C. Berglund. Copenhagen: World Health Organization.
- Yadav, M. *et al.* (2017) 'Auditory distraction in open-plan office environments: The effect of multi-talker acoustics', *Applied Acoustics*, 126, pp. 68–80.
- Yang, W. (2005) *An aesthetic approach to the soundscape of urban public open spaces.* PhD Thesis, School of Architecture, University of Sheffield, UK.
- Yang, W. and Kang, J. (2005) 'Soundscape and Sound Preferences in Urban Squares: A Case Study in Sheffield', *Journal of Urban Design*, 10(1), pp. 61–80.
- You, J., Lee, P. J. and Jeon, J. Y. (2010) 'Evaluating water sounds to improve the soundscape of urban areas affected by traffic noise', *Noise Control Eng. Journal*, 58(5), pp. 477–483.

Appendix A: Evaluation form used in the level preference test

SUBJECT X

Gender Male ☐ Female ☐

Age (years)

Nationality

Confirm you have no hearing difficulties (e.g. Tinnitus) by ticking this box ☐

PRACTICE TEST			
Test number	Sound 1 preferred	Sound 2 preferred	No preference
PRACTICE 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
PRACTICE 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
PRACTICE 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
PRACTICE 4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
PRACTICE 5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

LISTENING TESTS			
Test number	Sound 1 preferred	Sound 2 preferred	No preference
TEST 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 10	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 12	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 13	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 14	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 16	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 17	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 18	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 19	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 20	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

LISTENING TESTS			
Test number	Sound 1 preferred	Sound 2 preferred	No preference
TEST 21	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 22	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 23	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 24	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 25	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 26	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 27	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 28	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 29	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 30	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Test number	Sound 1 preferred	Sound 2 preferred	No preference
TEST 31	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 32	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 33	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 34	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 35	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 36	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 37	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 38	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 39	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 40	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 41	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 42	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 43	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 44	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 45	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 46	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 47	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 48	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 49	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TEST 50	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Appendix B: Instructions given to participants in the level preference test

SUBJECT X - INSTRUCTIONS

Imagine you are working in an open-plan office, where you can hear a **colleague speaking** from a nearby workstation. A **water feature** is also being played in the background and can be altered to make different sounds. In this listening test, you will hear **pairs of sound**, and you will have to **tick on the paper provided the sound which you prefer working in, over a long period of time** ('Sound 1', 'Sound 2' or 'No preference'). In your evaluation, take into account which sound might help you **concentrate better**.

PRACTICE TEST

As training, you will now listen to 5 paired comparisons, and you will tick your selection on the paper provided. Each paired comparison lasts 18 seconds. The listening tests are automated so **do not press any key while the test is ongoing**. **Although differences are subtle for some of the comparisons, always make every effort to indicate your preferred sound, rather than selecting 'No preference'.**

PRACTICE TEST

Put headphones on and press **ENTER** to start the practice test.

Practice 1...Practice 5.

PRACTICE TEST FINISHED

The practice test is now finished. **If you need any clarification about the procedure, please discuss with the tester.** You can request to repeat the practice test if necessary. Once you are confident with the procedure, please press **ENTER**.

LISTENING TESTS

You are now ready to start the listening tests. These include 50 paired comparisons, which correspond to a running time of approximately **25 minutes** (including **short breaks** between every 10 comparisons). Note that **once sound starts, you should not press any key**, as the tests are automated. **Please press ENTER to start the listening tests.**

Test 1... Test 10.

BREAK

If necessary, take a short break before continuing the test. If you want, you can also walk out of the room. **Once you are ready to continue the test, please press ENTER.**

Test 11... Test 20.

BREAK

Test 21... Test 30.

BREAK

Test 31... Test 40.

BREAK

Test 41... Test 50.

TEST FINISHED!!!!

THANK YOU FOR YOUR HELP AND PATIENCE!

Please hand over your marked sheet to the
tester and help yourself to more chocolates!

Appendix C: Evaluation form used in the audio-only and audio-visual preference and perception tests

SUBJECT X

Gender Male ☐ Female ☐

Age (years) Nationality

Confirm you have no hearing difficulties (e.g. Tinnitus) by ticking this box ☐

Section A

Practice tests

Please choose the option you prefer.

PART 1 - Audio-only test

Test number	Option 1 preferred	Option 2 Preferred
Practice 1	<input type="checkbox"/>	<input type="checkbox"/>
Practice 2	<input type="checkbox"/>	<input type="checkbox"/>

PART 2 - Audio-Visual test

Test number	Option 1 preferred	Option 2 Preferred
Practice 3	<input type="checkbox"/>	<input type="checkbox"/>
Practice 4	<input type="checkbox"/>	<input type="checkbox"/>

Actual tests

Please choose the option you prefer.

PART 1 - Audio-only test

Test number	Option 1 preferred	Option 2 Preferred
Test 1	<input type="checkbox"/>	<input type="checkbox"/>
Test 2	<input type="checkbox"/>	<input type="checkbox"/>
Test 3	<input type="checkbox"/>	<input type="checkbox"/>
Test 4	<input type="checkbox"/>	<input type="checkbox"/>
Test 5	<input type="checkbox"/>	<input type="checkbox"/>
Test 6	<input type="checkbox"/>	<input type="checkbox"/>
Test 7	<input type="checkbox"/>	<input type="checkbox"/>
Test 8	<input type="checkbox"/>	<input type="checkbox"/>
Test 9	<input type="checkbox"/>	<input type="checkbox"/>

Test number	Option 1 preferred	Option 2 Preferred
Test 10	<input type="checkbox"/>	<input type="checkbox"/>
Test 11	<input type="checkbox"/>	<input type="checkbox"/>
Test 12	<input type="checkbox"/>	<input type="checkbox"/>
Test 13	<input type="checkbox"/>	<input type="checkbox"/>
BREAK		
Test 14	<input type="checkbox"/>	<input type="checkbox"/>
Test 15	<input type="checkbox"/>	<input type="checkbox"/>
Test 16	<input type="checkbox"/>	<input type="checkbox"/>
Test 17	<input type="checkbox"/>	<input type="checkbox"/>
Test 18	<input type="checkbox"/>	<input type="checkbox"/>
Test 19	<input type="checkbox"/>	<input type="checkbox"/>
Test 20	<input type="checkbox"/>	<input type="checkbox"/>
Test 21	<input type="checkbox"/>	<input type="checkbox"/>
Test 22	<input type="checkbox"/>	<input type="checkbox"/>
Test 23	<input type="checkbox"/>	<input type="checkbox"/>
Test 24	<input type="checkbox"/>	<input type="checkbox"/>
Test 25	<input type="checkbox"/>	<input type="checkbox"/>

PART 2 - Audio-visual test

Test number	Option 1 preferred	Option 2 Preferred
Test 26	<input type="checkbox"/>	<input type="checkbox"/>
Test 27	<input type="checkbox"/>	<input type="checkbox"/>
Test 28	<input type="checkbox"/>	<input type="checkbox"/>
Test 29	<input type="checkbox"/>	<input type="checkbox"/>
Test 30	<input type="checkbox"/>	<input type="checkbox"/>
Test 31	<input type="checkbox"/>	<input type="checkbox"/>
Test 32	<input type="checkbox"/>	<input type="checkbox"/>
Test 33	<input type="checkbox"/>	<input type="checkbox"/>
Test 34	<input type="checkbox"/>	<input type="checkbox"/>
Test 35	<input type="checkbox"/>	<input type="checkbox"/>
Test 36	<input type="checkbox"/>	<input type="checkbox"/>
Test 37	<input type="checkbox"/>	<input type="checkbox"/>
Test 38	<input type="checkbox"/>	<input type="checkbox"/>
Test 39	<input type="checkbox"/>	<input type="checkbox"/>
Test 40	<input type="checkbox"/>	<input type="checkbox"/>

Section B

Practice tests

How your perception changed after introducing the water feature? Please tick one box.

PART 1 - Audio-only test

Test number	Much worse	Slightly worse	No change	Slightly better	Much better
Practice 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Practice 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

PART 2 - Audio-visual test

Test number	Much worse	Slightly worse	No change	Slightly better	Much better
Practice 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Practice 4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Actual tests

How your perception changed after introducing the water feature? Please tick one box.

Part 1 - Audio-only test

Test number	Much worse	Slightly worse	No change	Slightly better	Much better
Test 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test 4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test 5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test 6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Part 2 - Audio-visual test

Test number	Much worse	Slightly worse	No change	Slightly better	Much better
Test 7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test 8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test 9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test 10	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test 11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test 12	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

End of tests

Appendix D: Instructions given to participants the audio-only and audio-visual preference and perception tests

SUBJECT X – INSTRUCTIONS

Imagine you are working in an open-plan office, where you can hear **a colleague speaking** from a nearby workstation. A **water feature** is also being played in the background and can be altered to make different sounds. The test is divided into **two sections** (A and B), each consisting of **two parts**. Detailed instructions **about each part** are given **in the following slides**.

SECTION A

PART 1 – Audio-only

In this part of the test, you will **hear pairs of sounds**, and you will have to **tick on the paper provided the sound which you prefer working in, over a long period of time** (“Option 1”, or “Option 2”). In your evaluation, take into account which sound might help you **concentrate better**.

PART 2 – Audio-visual

This part is similar to Part 1, but **visual animations** are included. Therefore, both water sounds and their corresponding visual displays are used in the tests. Although the context is an occupied and busy office, the animations show an empty open-plan office, in order to focus the attention on the water feature displays. In the test, you will have to **tick on the paper provided the environment which you prefer working in, over a long period of time** (“Option 1”, or “Option 2”). In your evaluation, take into account which sound might help you **concentrate better**.

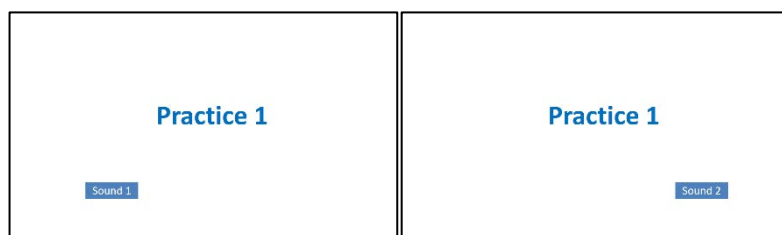
PRACTICE TEST

As training, you will now carry out **four practice tests**. You will need to **tick your selection on the paper provided**. Each test lasts 18 seconds. The tests are **automated**, so do not press any key while the test is ongoing.

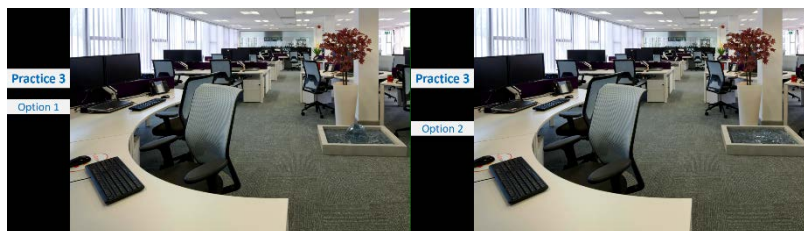
PRACTICE TEST (Section A)

Put headphones on and Press **ENTER** to start the practice test.

PART 1 – Audio-only (Practice 1 and Practice 2)



PART 2 – Audio-visual (Practice 3 and Practice 4)



PRACTICE TEST FINISHED

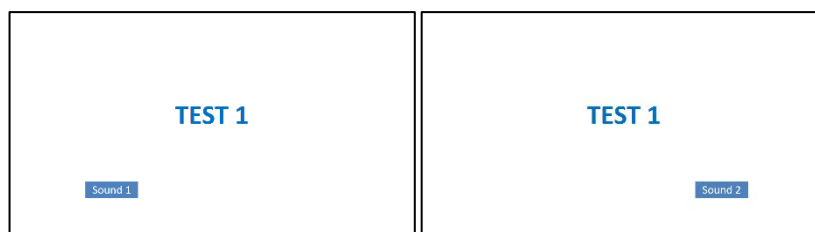
The practice test for **Section A** is now finished. **If you need any clarification about the procedure, please discuss with the tester.** You can request to repeat the practice test if necessary. Once you are confident with the procedure, please press **ENTER**.

ACTUAL TESTS

You are now ready to start the actual tests. These include 40 paired comparisons, corresponding to a running time of approximately **13 minutes** (including a **short break**). Note that **once the test starts, you should not press any key** unless you are asked to do so. **Please press ENTER to start the tests.**

ACTUAL TEST (Section A)

PART 1 – Audio-only (Test 1... Test 13).



BREAK

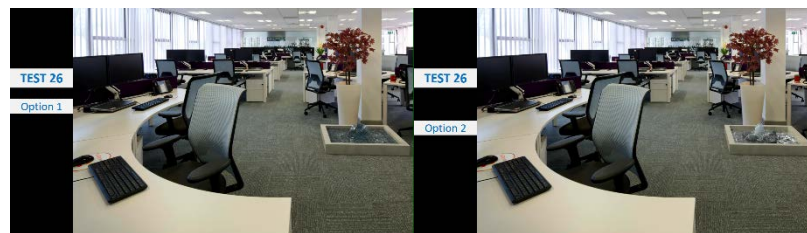
If necessary, take a short break before continuing the test. If you want, you can also walk out of the room. **Once you are ready to continue the test, please press ENTER.**

(Test 14... Test 25, same as above)

BREAK

ACTUAL TEST (Section A)

PART 2 - Audio-visual (Test 26... Test 40).



SECTION A FINISHED

Section A of the test is finished. Please **call the tester in** to give you instructions on Section B of the test.

SECTION B

PART 1 – Audio-only

In this part, you will hear 7 seconds of background noise of an office, which includes **audible speech from a nearby colleague**. A **water sound** will then be added to the background noise for comparison. In the test, you should **state how your perception changed after the water sound was added to the background noise**: “*Much worse, Slightly worse, No change, Slightly better, and Much better*”. In your evaluation, take into account which sound might help you **concentrate better**.

PART 2 – Audio-visual

This part of the test is similar to part 1 but **visual animations** are included. Therefore, both **water sounds and their corresponding visual displays** are used in the tests. Similar to Part 1, you should **state how your perception of the environment changed with the inclusion of a water feature**: “*Much worse, slightly worse, No change, slightly better, and Much better*”. In your evaluation, take into account which sound might help you **concentrate better**.

PRACTICE TEST

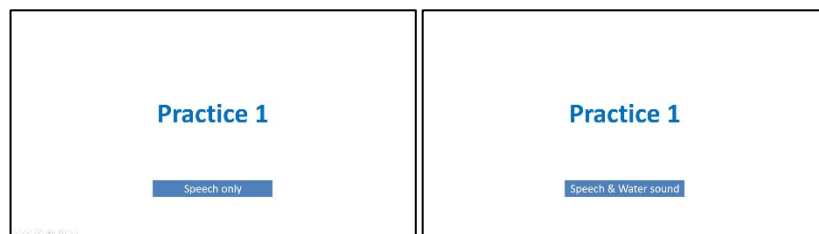
As training, you will now carry out **4 practice tests**. You will need **to tick your selection on the paper provided**. Each test lasts 20 seconds. The tests are **automated**, so do not press any key while the test is ongoing.

PRACTICE TEST (Section B)

Put headphones on and press **ENTER** to start the practice test.

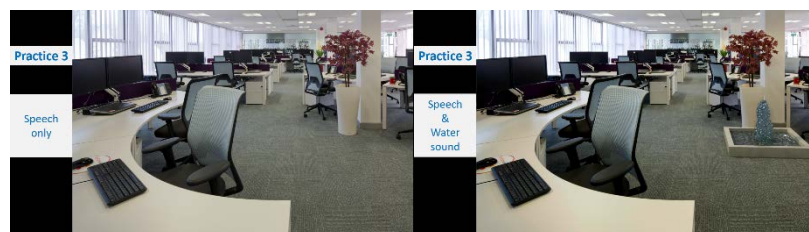
PRACTICE TEST (Section B)

PART 1 – Audio-only (Practice 1 and Practice 2)



PRACTICE TEST (Section B)

PART 2 – Audio-visual (Practice 3 and Practice 4)



PRACTICE TEST FINISHED

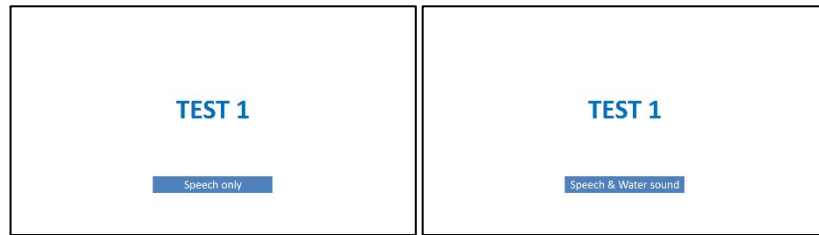
The practice test for **Section B** is now finished. **If you need any clarification about the procedure, please discuss with the tester.** You can request to repeat the practice test if necessary. Once you are confident with the procedure, please press **ENTER**.

ACTUAL TESTS

You are now ready to start the actual tests. These include 12 questions about your perception, corresponding to a running time of approximately **5 minutes** (including a short break). Note that **once the test starts, you should not press any key** unless you are asked to do so. **Please press ENTER to start the tests.**

ACTUAL TEST

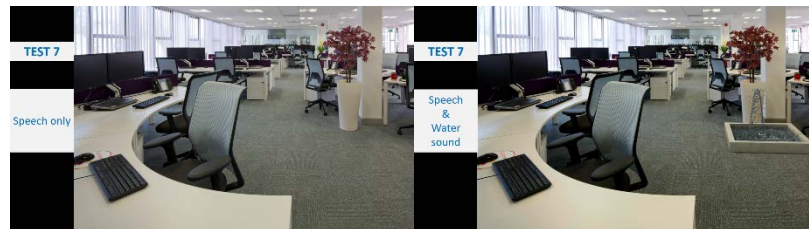
PART 1 – Audio-only (Test 1... Test 6)



BREAK

ACTUAL TEST

PART 2 - Audio-visual (Test 7... Test 12)



TEST FINISHED!!!!

THANK YOU FOR YOUR HELP AND PATIENCE!

Please hand over your marked sheet to the tester and help yourself to more chocolates!

Appendix E: General instructions for the cognitive performance tasks

This experiment will measure your ability in performing certain tasks under different background noise conditions. First, you will fill in a short questionnaire on background information (e.g. age, gender etc.), followed by another questionnaire on your sensitivity to noise. A practice session will then follow, illustrating each of the tasks making up the experiment. When you are ready please click “NEXT”.

- Please confirm that you do not have any known hearing difficulties (e.g. Tinnitus), or vision impairment (e.g. cataracts).
- Please confirm that you do not have any known learning difficulties (e.g. Dyslexia)
- What is your gender?
- What is your nationality?
- What is your age?
- How many hours did you sleep last night?

Noise Sensitivity Questionnaire

The following questionnaire consists of **4** statements which relate to your typical response to sounds in working spaces. Please indicate your level of agreement or disagreement for each of the statements. There are no right or wrong answers. We simply would like to know your opinions. Please click “NEXT” to see the statements.

1. I need a quiet environment to be able to carry out new tasks.
2. When people around me are noisy, I find it hard to do my work.
3. I perform significantly worse in noisy environments.
4. I need peace and quiet in order to carry out a difficult task.

Strongly disagree

☐

Slightly disagree

☐

Slightly agree

☐

Strongly agree

☐

Noise Sensitivity Questionnaire Finished!!

Thanks for completing the questionnaire. Please click “NEXT” to start the practice session.

Instructions

Imagine that you are working in an open-plan office, where you can hear a colleague speaking from a nearby workstation. You are required to carry out certain tasks under **two different background noise conditions**. In one case, you will only hear the **speech by your colleague**, while in the other, you will hear the **speech and a water sound**. In each case, you are required to carry out a set of **4 tasks**. You should try to ignore the background noise and **concentrate on the tasks**. No questions will be asked on the content of the background speech.

Practice session

As practice, you will now carry out all of the 4 tasks, in order to get used to them. The scores in the practice tests will not be analysed and are therefore irrelevant. You can repeat the practice tests as many times as you wish, until you are confident to carry out the actual tasks. If you are ready, please click “NEXT” to start the first practice test.

Practice 1: Serial recall task

Instructions given in Appendix F.

Serial recall practice task Finished!

You have finished the practice test for this task. If you would like to repeat this practice test, please click on “AGAIN”, otherwise, please click “NEXT” to proceed to the next practice test.

Practice 2: Information matching task

Instructions given in Appendix G.

Information matching practice task Finished!

You have finished the practice test for this task. If you would like to repeat this practice test, please click on “AGAIN”, otherwise, please click “NEXT” to proceed to the next practice test.

Practice 3: One-back task

Instructions given in Appendix H.

One-back practice task Finished!

You have finished the practice test for this task. If you would like to repeat this practice test, please click on AGAIN, otherwise, please click NEXT to proceed to the last practice test.

Practice 4: Reading comprehension task

Instructions given in Appendix I.

Reading comprehension practice task Finished!

You have finished the last part of the practice tests. If you would like to repeat this last task, please click on “AGAIN”. If you have any question about the experiment, or you would like to address any issue, please ask the tester. Once you are ready to continue and start the actual tests, please click “NEXT”.

Practice tasks finished!

You have now successfully completed the practice tasks. If there is someone else in the room taking the test with you, please allow them to complete the tests. Once everyone in the room has completed the practice tasks, you may proceed with the actual tests by clicking on the button below.

Actual Tests: Session one

(Appendices F, G, H, and I)

Satisfaction Questionnaire

The following questionnaire consists of 9 statements which relate to your perception of the sound environment, based on your experience in the last few minutes. Please indicate your level of agreement or disagreement for each of the statements. There are no right or wrong answers. We simply would like to know your opinions. Please click “NEXT” to see the statements.

1. The sound environment was pleasant.
2. My attention was drawn to the sound environment.
3. The sound environment helped me to concentrate on the tasks.
4. The speech disturbed me.

For the next statements, please think what it would be like if you had to work all the time in conditions similar to that experienced in the last few minutes.

5. I could have meetings in my office without distracting others.
6. I could work uninterrupted for long periods.
7. The noise in my office would not be distracting.
8. I could easily have confidential conversations.

Strongly disagree								Strongly agree		
0	1	2	3	4	5	6	7	8	9	10
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

9. What percentage of the speech did you understand? (%)

Thank you!

The tasks and questionnaire for this part of the experiment is now finished. Kindly complete the tests by clicking on “FINISH” and wait for the tester to provide you with further instructions. If there is someone else in the room taking the test with you, please allow them to complete before calling the tester.

Actual Tests: Session one

(Appendices F, G, H, and I)

Satisfaction Questionnaire

(same as above)

End of experiment!

Kindly complete the experiment by clicking on “FINISH”. Your time and effort are greatly appreciated.

Appendix F: Serial recall task and instructions

In the next slides, you will see **nine** digits from 1 to 9. These will be presented in a specific order, with one digit per slide. After you have seen all the nine digits, a keypad will appear on the screen: you should then type the digits in the order in which they were presented. If you cannot remember a digit, you can click on the “×” button.

There will be **FIVE** trials of this task. Each trial will last for around **30 seconds**. If you would like to repeat the practice test for this task, please click on “Practice again”. If you do not need the practice task to be repeated, please click on “**START**” to start the first trial of the actual task.

Trial	Condition 1	Condition 2
1	4 1 5 8 2 7 6 3 9	5 3 6 4 2 1 7 9 8
2	6 3 5 4 9 7 1 8 2	9 8 2 1 6 4 5 7 3
3	6 2 1 8 4 5 3 9 7	1 4 2 8 5 6 9 7 3
4	5 1 3 2 8 4 7 9 6	3 9 4 2 5 1 7 6 8
5	1 4 8 3 7 2 6 5 9	5 7 3 6 2 9 8 1 4

1

2

3

4

5

6

7

8

9

×

←

NEXT➤

Serial recall task finished!

On a scale of 0 to 10, how difficult the task was?

Extremely easy 1 2 3 4 5 6 7 8 9 10 Extremely difficult

Appendix G: One-back task and instructions

In this task, you will be presented with a list of letters. Each time you see a letter, you will need to decide if **the letter is the same as the one you saw on the previous slide** (click on ✓), or if the letter is different (click on ✕).

There will be **ONE** trial of the task. The task will last for around **90 seconds**. If you would like to repeat the practice test for this task, please click on “Practice again”. If you do not need the practice task to be repeated, please click on “**START**” to start the actual task.

Con.1	d	U	c	C	j	J	L	v	R	r	f	o	O	b	B	P	w	W	G	g	q	h	H	y	Y	e	n	S	s	M	k
Con.2	h	z	n	N	v	W	w	s	q	m	M	D	c	y	Y	o	O	L	f	k	P	p	A	a	X	E	e	B	b	j	J

One-Back task finished!

On a scale of 0 to 10, how difficult the task was?

Extremely easy 1 2 3 4 5 6 7 8 9 10 Extremely difficult

Appendix H: Information matching task and instructions

In the next slide, you will see a **table** on the left-hand side of the screen. The table contains nine rows ordered from 1 to 9, where each row consists of a **letter**, a **digit** (1 or 2), and a **colour** (B or orange). You should then fill in the **3×3 matrix** by clicking on the small squares of the on-screen keyboard, where **1 shown in the matrix should match with row 1 of the table, 2 should match with row 2** etc.

There will be **FIVE** trials of the task. Each trial will last for around **30 seconds**. If you would like to repeat the practice test for this task, please click on “Practice again”. If you do not need the practice task to be repeated, please click on “**START**” to start the first trial of the actual task.

	Trial 1			Trial 2			Trial 3			Trial 4			Trial 5			
Order	Letter	Number	Colour*	Letter	Number	Colour*	Letter	Number	Colour*	Letter	Number	Colour*	Letter	Number	Colour*	
Condition 1	1	H	2	B	D	1	Y	G	2	B	F	1	B	B	2	Y
	2	B	2	Y	G	2	B	E	1	Y	C	2	B	F	1	Y
	3	D	1	Y	H	1	Y	A	1	B	I	2	Y	A	2	Y
	4	G	2	B	B	1	Y	H	1	B	E	2	B	I	1	B
	5	F	2	Y	F	2	Y	D	2	B	D	2	B	C	2	B
	6	C	2	Y	E	2	B	B	1	B	H	2	Y	D	2	B
	7	E	2	B	C	1	B	F	2	Y	G	1	Y	H	1	Y
	8	A	2	Y	I	1	B	I	1	Y	A	2	B	G	2	Y
	9	I	1	B	A	1	B	C	1	B	B	2	Y	E	2	Y
Condition 1	1	H	1	B	H	2	Y	D	1	Y	D	2	Y	B	2	Y
	2	E	1	Y	B	1	B	I	1	B	F	1	Y	I	2	B
	3	D	2	B	F	2	Y	B	1	Y	G	2	B	A	1	Y
	4	G	2	Y	I	1	Y	E	1	Y	C	2	Y	D	1	Y
	5	A	2	B	A	1	B	F	2	Y	E	1	B	G	2	Y
	6	C	1	Y	E	1	B	G	1	Y	A	2	Y	E	2	B
	7	F	1	Y	D	1	B	H	1	B	I	2	Y	C	2	B
	8	B	2	B	C	2	B	C	2	B	H	1	B	H	1	B
	9	I	1	B	G	1	Y	A	1	Y	B	1	Y	F	2	B

*B = Blue. Y = Yellow.

Information matching task finished!

On a scale of 0 to 10, how difficult the task was?

Extremely easy 1 2 3 4 5 6 7 8 9 10 Extremely difficult

Appendix I: Reading comprehension task and instructions

In the following slides, you will see a **three-sentence paragraph** which describes the relationship between five things. Two of the things are real words which have a meaning, while the remaining three are nonsense words invented for the purpose of this task. Each sentence will be presented separately and should be read **very carefully**. After you have read the three sentences, you will be presented with **statements** and you will need to indicate if they are **true** (click on ✓), or **false** (click on ✕).

Your decisions should be based on the information presented to you in the sentences as well as your own knowledge. It is **crucial to read the sentences very carefully**. The words marked in **RED** are nonsense words which have no meaning in the English language.

There will be **One** trial of the task. The task will last for around **5 minutes**. If you would like to repeat the practice test for this task, please click on “Practice again”. If you do not need the practice test to be repeated, please click on “**START**” to start the first trial of the actual task.

Condition 1

A **MIRT** resembles a Lion but is larger and eats more.

A **COFT** resembles a Chicken but is smaller and eats more.

A **FILP** resembles a **COFT** but is smaller and eats more.

Type of task	Level	True	False
Memory	N/A	1 A MIRT is larger than a Lion.	A lion is larger than a MIRT.
		2 A COFT is smaller than a Chicken.	A Chicken is smaller than a COFT.
		3 A FILP is smaller than a COFT.	A COFT is smaller than a FILP.
		4 A MIRT eats more than a Lion.	A Lion eats more than a MIRT.
		5 A COFT eats more than a chicken.	A Chicken eats more than a COFT.
		6 A FILP eats more than a COFT.	A COFT eats more than a FILP.
Inferencing	N/A	7 A FILP is smaller than a Chicken.	A Chicken is smaller than a FILP.
		8 A FILP eats more than a Chicken.	A Chicken eats more than a FILP.
Knowledge Integration	Low	9 A MIRT eats more than a Chicken.	A Chicken eats more than a MIRT.
		10 A Lion is larger than a COFT.	A COFT is larger than a LION.
	Mid	11 A MIRT is larger than a Duck.	A duck is larger than a MIRT.
		12 A Wolf is larger than a COFT.	A COFT is larger than a Wolf.
	high	13 A Wolf is larger than a FILP.	A FILP is larger than a Wolf.
		14 Like Wolves MIRTs can't lay eggs.	Like Ducks, MIRTs can lay eggs.
Knowledge Access	Low	15 Like Ducks, COFTs can lay eggs.	Like Wolves, COFTs can't lay eggs.
		16 Like Ducks, FILPs can lay eggs.	Like Wolves, FILPs can't lay eggs.
	High	17 A Lion is larger than a Chicken.	A Chicken is larger than a Lion.
		18 A Lion eats more than a Chicken.	A Chicken eats more than a Lion.
	High	19 A Chicken can be domesticated, whereas a Wolf typically can't	A Wolf can be domesticated, whereas a Chicken typically can't
		20 A Duck can be domesticated whereas a Lion typically can't	A Lion can be domesticated, whereas a Duck typically can't

Condition 2

A **TOLP** resembles a **MARB** but is more colourful and larger.

A **MARB** resembles a Butterfly but is more colourful and larger.

A **JERP** resembles an Ant but is less colourful and larger.

Type of task	Level	True	False
Memory	N/A	1 A TOLP is more colourful than a MARB	A MARB is more colourful than a TOLP
		2 A MARP is more colourful than a Butterfly	A Butterfly is more colourful than a MARB
		3 A JERP is less colourful than an Ant	An Ant is less colourful than a JERP
		4 A TOLP is larger than a MARB	A MARB is larger than a TOLP
		5 A MARB is larger than a BUTTERFLY	A Butterfly is larger than a MARB
		6 A JERP is larger than an ANT	An Ant is larger than a JERP
Inferencing	N/A	7 A TOLP is more colourful than a Butterfly	A Butterfly is more colourful than a TOLP
		8 A TOLP is Larger than a Butterfly	A Butterfly is larger than a TOLP
Knowledge Integration	Low	9 A MARB is larger than an Ant	An Ant is larger than a MARB
		10 A Butterfly is more colourful than a JERP	A JERP is more colourful than a Butterfly
	Mid	11 A TOLP is more colourful than a Spider	A Spider is more colourful than a TOLP
		12 A MARB is more colourful than a Spider	A Spider is more colourful than a MARB
		13 A Bee is more colourful than a JERP	A JERP is more colourful than a Bee
	high	14 Like Bees, TOLPs can fly	Like Spiders, TOLPs can't fly
		15 Like Bees, MARBs can fly	Like Spiders, MARBs can't fly
		16 Like Spiders, JERPs can't fly	Like Bees, JERPs can fly
Knowledge Access	Low	17 A Butterfly is more colourful than an Ant	An Ant is more colourful than a Butterfly
		18 A Butterfly is larger than an Ant	An Ant is Larger than a Butterfly
	High	19 Ants have queens, whereas Spiders don't	Spiders have queens whereas Ants don't
		20 Bees have queens, whereas Butterflies don't	Butterflies have queens, whereas Bees don't

Reading Comprehension task finished!

On a scale of 0 to 10, how difficult the task was?

Extremely easy 1 2 3 4 5 6 7 8 9 10 Extremely difficult

Appendix J: Satisfaction questionnaire used in Experiment 4

(Part 1)

The aim of this questionnaire is to obtain your views on how noise affects you in your work environment. All answers provided will be treated in the strictest confidence. There is no right or wrong answer. Please just give your honest opinion.

This is the first part of the questionnaire. The second part will be distributed in about two weeks from now. ***Please keep this with you until the second part of the questionnaire is distributed.***

General information about you and your work station

Please tick the box that corresponds to your situation and enter the relevant information

1. Gender: male ☐ female ☐
2. Age: years
3. Do you have any known hearing disabilities?
Yes ☐ No ☐

If you answered "yes", please specify.....

4. Length of time working in this office: years
5. Is your work station allocated on a permanent basis in the office? Yes ☐ No ☐

If you answered "No", please state how many days in a week you are in this office.....

Assessing the physical environment of your work area

The following statements concern your physical work environment. For each statement, please respond by circling the number that corresponds to your level of satisfaction on a scale of 0 to 10 where zero is "very dissatisfied" and 10 is "very satisfied".

6. Noise environment (Scale 1).
7. Possibility of concentrating in your workplace (Scale 1).
8. The quality of the lighting (Scale 1).

9. The physical position of your work station (Scale 1).
10. Possibility of having private conversations (Scale 1).
11. Possibility of managing noise (Scale 1).
12. The furniture in your work area (Scale 1).
13. Possibility of seeing outside (Scale 1).
14. The cleanliness of your work area (Scale 1).
15. The equipment available in your work area (Scale 1).
16. Possibility of controlling the temperature (Scale 1).
17. The air circulation in your work area (Scale 1).
18. Possibility of personalizing your work area (with personal objects, photos...) (Scale 1).
19. Possibility of not being seen by others (Scale 1).

Assessing the noise environment of your work area

The following questions concern noise in your work environment. For each question, please respond by circling the number that corresponds to your level of agreement on a scale of 0 to 10 where zero is "not at all/never" and 10 is "extremely/constantly".

20. Generally speaking, would you say the level of noise in your work environment is high? (Scale 2)
21. Generally speaking, would you say the noise in your work environment bothers you? (Scale 3)
22. At your work station, you clearly hear and **CAN** understand your colleagues' conversations: (Scale 3)

☞ If you circled "0", go to question 26.
23. Would you say that these conversations bother you? (Scale 2)

☞ If you circled "0", go to question 26.
24. Is there a task in your work for which these conversations are particularly distracting? Yes ☐ No ☐

If you answered "yes", which task? (For example, reading, writing, data entry, telephone conversations, etc.)

25. It bothers you most when:

You can hear all of the speakers ☐
(e.g. conversations of people in the office)

You can hear only one speaker ☐
(e.g. telephone conversations)

You are equally bothered by both of the above ☐

26. At your work station, you hear colleagues' conversations that you **CANNOT** understand: (Scale 3)

☞ If you circled "0", go to question 30.

27. Would you say that these conversations bother you? (Scale 2)

☞ If you circled "0", go to question 30.

28. Is there a task in your work for which these conversations are particularly distracting? Yes ☐ No ☐

29. If you answered "yes", which task? (For example, reading, writing, data entry, telephone conversations, etc.).....

30. For you, the ideal work environment would be:

An open-plan office ☐

A private office ☐

A shared office, (shared with 1 - 3 people) ☐

Other (please state) ☐

31. Please **order** the following noise sources, from the most distracting to the least distracting. Please tick number 1 to indicate the element you find the most distracting, number 2 to indicate the element which is the next most distracting and so on until you reach the least distracting element (8).

- Ventilation/Air-conditioning noise.
- Printers/Photocopiers.
- Telephones ringing.
- Conversations which you can hear but cannot understand.
- Conversations in which you can understand what is said.
- People walking up and down the office.
- Noise of people working (keyboard, opening and closing drawers, etc.)
- Noise linked to one person in particular

Your perception of the sound environment

The following statements relate to your perception of the sound environment in your office. For each question, please respond by circling the number that corresponds to your level of agreement on a scale of 0 to 10 where zero is "strongly disagree" and 10 is "strongly agree".

32. The sound environment is pleasant (Scale 4).
33. The sound environment helps me to concentrate on my tasks (Scale 4).
34. I can have meetings in my office without distracting others (Scale 4).
35. I can work uninterrupted for long periods (Scale 4).
36. The noise in my office is distracting (Scale 4).
37. I can have confidential conversations at my work station (Scale 4).
38. The sound environment helps me to concentrate on my tasks (Scale 4).

Your relationship with noise in general (at home, at night and at work)

Please circle the number that corresponds to your level of agreement with the following statements.

39. I need a completely quiet environment in order to have a good night's sleep (*Sleep*) (Scale 5).
40. I need a quiet environment to be able to carry out new tasks (*Work*) (Scale 5).

41. When I am at home, I quickly get used to noise (*Habitation*) (Scale 5).
42. I get very agitated when I hear somebody speaking when I am trying to sleep (*Sleep*) (Scale 5).
43. I am very sensitive to noise made by the neighbours (*Habitation*) (Scale 5).
44. When people around me are noisy, I find it hard to do my work (*Work*) (Scale 5).
45. I perform significantly worse in noisy environments (*Work*) (Scale 5).
46. I do not feel very alert when I have been bothered by noise the night before (*Sleep*) (Scale 5).
47. Living in a noisy street would not bother me (*Habitation*) (Scale 5).
48. I am prepared to accept disadvantages in order to live in a quiet place (*Habitation*) (Scale 5).
49. I need peace and quiet in order to carry out a difficult task (*Work*) (Scale 5).
50. I can sleep even if it is noisy (*Sleep*) (Scale 5).

Thank you for participating

Scales

Scale 1	Very dissatisfied	0	1	2	3	4	5	6	7	8	9	10	Very satisfied
Scale 2	Not at all	0	1	2	3	4	5	6	7	8	9	10	Extremely
Scale 3	Never	0	1	2	3	4	5	6	7	8	9	10	Constantly
Scale 4	Strongly disagree	0	1	2	3	4	5	6	7	8	9	10	Strongly agree
Scale 5	Completely disagree	1				2			3			4	Completely agree

Appendix K: Satisfaction questionnaire used in Experiment 4

(Part2)

This is the second part of a questionnaire distributed around two weeks ago. This part of the questionnaire aims at obtaining your views on how noise affects you in your work environment. The questionnaires also assess the effectiveness of a water feature placed in your office for improving your work environment over the last two weeks.

Most parts of this questionnaire are identical to those answered in the first part. If you want, you can compare your responses in this part of the questionnaire to those given in the previous part. Duplicate questions/statements are marked by a U sign followed by a number. The number represents the order of the questions/statements in the first part of the questionnaire to allow for easy comparisons.

Assessing the physical environment of your work area

Same as in Appendix I.

Assessing the noise environment of your work area

Same as in Appendix I.

Your perception of the sound environment

Same as in Appendix I, plus the following Items:

- 39. The water sound is pleasant.
- 40. The water feature has improved the sound environment.
- 41. The water feature helps me to carry out private conversations.
- 42. The water feature stresses me.
- 43. The water sound is distracting.
- 44. The water feature is visually/aesthetically pleasing.
- 45. Does the water feature help you to carry out a particular task in a better way?

Yes ☐ No ☐. If you answered "yes", which task?

46. Does the water feature make you go to the toilet more frequently?
- Yes ☐ No ☐. If you answered "**yes**", please state by how many time(s) per day?
47. Would you like the water feature to remain in your office permanently?
- Yes ☐ No ☐
48. If you would like to address anything else related to the water feature or the sound it makes, please do so using the space below.

Thank you for participating!

Appendix L: Participants' response to satisfaction questionnaire in Experiment 4

Table L.1 Participants responses to the satisfaction questionnaire before installing the water feature.

Participant		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	
Background information	Gender	F	F	F	F	M	F	F	F	F	F	F	F	M	F	Av.
	Age (yr)	52	36	24	35	25	25	25	47	45	49	51	47	36	61	39.86
	Disability	No	No	No	No	No	No	No	No	No	No	Yes†	No	No	No	N/A
	Length of time (yr)	1.5	1	1	0.5	1	0.5	1	1	5	1	1	1	3	1	1.39
	Day/Week	5	1	2	1	3	2	3	3	3	3	1	5	5	1	5
Physical Environment	Q6*	9	7	6	7	8	9	8	5	8	7	3	7	5	8	6.93
	Q7**	6	6	8	6	7	9	8	9	6	9	3	7	6	8	7.00
	Q8*	9	10	6	7	10	10	8	9	9	10	9	3	5	7	8.00
	Q9*	10	8	4	6	4	10	6	10	10	10	8	4	4	8	7.29
	Q10**	5	5	6	5	3	9	6	2	8	1	8	6	4	9	5.50
	Q11**	6	5	8	6	6	9	7	5	6	1	2	2	4	9	5.43
	Q12*	7	8	8	6	6	9	9	9	8	10	7	8	5	7	7.64
	Q13**	9	9	8	8	10	8	8	7	10	4	8	10	10	9	8.43
	Q14*	7	6	8	8	10	9	9	7	10	2	8	9	9	9	7.93
	Q15*	8	6	8	8	10	10	9	7	10	10	8	9	10	9	8.71
	Q16**	6	5	4	5	4	8	8	6	8	5	8	5	5	9	6.14
	Q17*	6	5	5	5	8	8	8	6	10	5	8	5	4	9	6.57
	Q18**	8	6	5	3	8	9	9	9	10	0	8	8	6	9	7.00
	Q19**	9	3	5	3	3	9	5	5	10	10	5	6	5	9	6.21
Noise Environment	Q20	7	4	3	5	7	4	2	0	5	2	9	7	3	4	4.43
	Q21	8	2	2	5	3	2	5	4	5	1	8	7	4	2	4.14
	Q22	7	8	6	8	5	7	9	5	8	10	10	7	6	9	7.50
	Q23	8	3	4	6	2	1	2	5	3	0	8	8	4	1	3.93
	Q24															
	Q24Part2	See Table L.2														
	Q25															
	Q26	8	1	0	4	8	7	8	5	3	0	2	6	1	0	3.79

Table L.1 Participants responses to the satisfaction questionnaire before installing the water feature (continued).

Participant		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	
	Q27	8	2	0	5	0	1	1	5	2	0	2	7	1	0	2.43
	Q28															
	Q29	See Table L.2														
	Q30															
	Q31(1)	2	7	8	5	7	7	7.5	-	8	1	8	8	8	-	6.38
	Q31(2)	4	6	6	8	6	1	1	-	6	7	7	7	4	-	5.25
	Q31(3)	7	5	1	3	1	2	2	-	1	2	4	4	2	-	2.83
	Q31(4)	8	2	5	2	4	4	3	-	3	5	6	2	5	-	4.08
	Q31(5)	6	1	4	1	3	3	4	-	2	6	1	3	1	-	2.92
	Q31(6)	3	3	7	7	2	6	5	-	5	4	2	5	6	-	4.58
	Q31(7)	5	4	2	6	5	5	6	-	7	3	5	6	3	-	4.75
	Q31(8)	1	8	3	4	8	8	7.5	-	4	8	3	1	7	-	5.21
Perception of Sound Environment	Q32	9	7	6	6	4	8	3	7	7	8	2	6	6	8	6.21
	Q33	8	6	5	3	5	7	4	8	5	9	2	4	5	2	5.21
	Q34	3	1	5	3	2	2	4	2	2	2	7	4	4	9	3.57
	Q35	3	4	8	6	3	8	9	7	0	2	2	6	0	9	4.79
	Q36	3	4	4	7	4	2	2	5	3	3	9	6	5	8	4.64
	Q37	1	0	5	0	0	3	1	0	0	0	2	4	0	9	1.79
	Q38	4	4	5	3	5	8	4	6	3	5	2	5	5	3	4.43
Your relationship with noise in general	Q39	3	3	2	4	4	1	1	4	1	1	4	4	2	4	2.71
	Q40	4	3	4	3	3	2	1	3	2	2	3	4	2	2	2.71
	Q41	3	4	3	3	3	3	4	4	4	4	2	4	3	4	3.43
	Q42	4	2	2	3	4	2	2	4	2	4	4	4	3	2	3.00
	Q43	4	2	1	2	1	2	2	4	1	1	3	2	2	3	2.14
	Q44	3	3	2	3	3	2	1	3	3	2	3	3	3	2	2.57
	Q45	3	3	2	3	3	2	1	3	2	2	4	3	2	4	2.64
	Q46	4	3	3	3	4	2	1	3	1	3	4	3	3	2	2.79
	Q47	1	2	2	2	1	3	4	1	4	3	1	1	2	2	2.07
	Q48	3	2	2	2	4	2	1	3	1	4	4	3	2	3	2.57
	Q49	3	3	3	3	3	2	1	3	3	2	4	3	2	3	2.71
	Q50	1	2	4	3	1	3	4	2	4	4	1	1	2	1	2.36

*Comfort/Functionality Subscale. **Control/Privacy subscale. † Hard of hearing- Right side

Table L.2 Participants responses to Questions 24, 25, 28, 29, and 30 in the satisfaction questionnaire before installing the water feature.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Q24	No	No	Yes	Yes	Yes	No	No	No	Yes	No	Yes	Yes	Yes	No
Q24 Part 2	N/A	N/A	Writing on important spreadsheets	Writing	HR record updating	N/A	N/A	N/A	Report writing	N/A	Reading and writing	Telephone conversation, data entry	Telephone Conversati on	N/A
Q25	Equally bothered by both	Hear only one speaker	Equally bothered by both	Equally bothered by both	Hear all speakers	Hear all speaker s	Equally bothered by both	Hear only one speaker	Hear only one speaker	Equally bothered by both	Equally bothered by both	Equally bothered by both	Equally bothered by both	Hear all speaker s
Q28	Yes	No	No	Yes	No	No	No	No	Yes	No	No	Yes	No	No
Q29	Payroll entry	N/A	N/A	Writing	N/A	N/A	N/A	N/A	Report writing	N/A	N/A	Telephone Conversation, data entry	N/A	N/A
Q30	Shared office	Open- Plan office	Open-Plan office	Private office	Private office	Open- Plan office	Other	Shared office	Open- Plan office	Open-Plan office	Shared office	Shared office	Other	Shared office

Table L.3 Participants responses to the satisfaction questionnaire after installing the water feature.

Participant		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	
Physical Environment	Q6*	9	8	9	8	9	8	8	7	9	9	4	8	6	10	8.00
	Q7**	7	6	9	8	10	9	8	7	7	9	4	8	6	10	7.71
	Q8*	9	8	7	6	10	10	8	8	9	10	9	6	5	10	8.21
	Q9*	10	7	2	7	4	10	6	10	10	10	9	4	4	10	7.36
	Q10**	5	2	6	5	3	9	6	0	8	1	8	5	4	10	5.14
	Q11**	6	4	7	5	6	9	7	5	6	1	4	3	4	12	5.64
	Q12*	7	7	8	7	6	9	9	8	8	9	9	5	5	6	7.36
	Q13**	9	9	8	9	10	9	8	6	10	3	10	10	10	10	8.64
	Q14*	7	9	8	8	10	9	9	9	10	4	9	9	9	10	8.57
	Q15*	8	9	8	7	10	10	9	8	10	7	9	9	10	8	8.71
	Q16**	6	5	7	5	4	8	8	5	8	5	8	4	5	10	6.29
	Q17*	6	5	7	5	8	8	8	5	10	5	8	4	4	10	6.64
	Q18**	8	6	7	5	8	9	9	9	10	0	8	5	6	10	7.14
	Q19**	9	3	6	3	3	9	5	5	10	8	8	8	5	10	6.57
Noise Environment	Q20	7	4	6	5	7	2	2	1	6	0	8	4	3	5	4.29
	Q21	7	5	5	7	2	2	5	5	6	0	7	7	3	0	4.36
	Q22	6	7	5	7	5	8	9	5	7	4	8	7	6	3	6.21
	Q23	7	4	4	6	2	1	2	5	3	2	6	8	4	0	3.86
	Q24															
	Q24Part2								See Table L.4							
	Q25															
	Q26	7	6	5	3	8	8	8	5	3	5	0	5	1	0	4.57
	Q27	7	5	5	5	0	1	1	5	3	8	0	8	1	0	3.50
	Q28															
	Q29								See Table L.4							
	Q30															
	Q31(1)	2	7	8	5	7	7	7.5	-	8	1	7	8	8	-	6.29
	Q31(2)	4	6	2	7	6	1	1	-	6	7	5	7	4	-	4.67
	Q31(3)	7	3	5	4	1	2	2	-	1	2	3	6	2	-	3.17
	Q31(4)	8	2	6	2	4	4	3	-	3	5	8	3	5	-	4.42
	Q31(5)	5	1	1	1	3	3	4	-	2	6	1	4	1	-	2.67

Table L.3 Participants responses to the satisfaction questionnaire after installing the water feature (continued)

Participant		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	
	Q31(6)	6	4	7	3	2	6	5	-	5	4	2	2	6	-	4.33
	Q31(7)	3	5	3	6	5	5	6	-	7	3	6	5	3	-	4.75
	Q31(8)	1	8	4	8	8	8	7.5	-	4	8	4	1	7	-	5.71
Perception of Sound Environment	Q32	10	8	8	6	7	9	6	10	8	10	3	7	7	10	7.79
	Q33	9	7	8	6	6	8	6	6	5	10	3	6	6	10	6.86
	Q34	6	4	4	3	2	2	4	6	2	6	3	4	4	10	4.29
	Q35	3	5	7	5	3	8	9	8	0	9	1	7	0	10	5.36
	Q36	3	3	5	6	3	1	2	0	2	0	6	6	5	0	3.00
	Q37	1	2	4	3	0	2	1	1	0	8	2	3	0	2	2.07
	Q38	7	5	6	6	6	9	6	5	2	10	4	6	6	10	6.29
Question specific to the water feature	Q39	10	10	9	7	10	10	9	10	8	10	8	9	8	10	9.14
	Q40	9	8	9	8	10	9	8	10	5	10	8	8	8	10	8.57
	Q41	6	2	9	4	4	9	4	3	0	10	4	2	0	5	4.43
	Q42	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0.14
	Q43	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0.14
	Q44	10	10	9	8	10	10	10	10	10	10	9	9	9	10	9.57
	Q45	No	No	No	Yes	No	No	No	No	No	Yes	Yes	No	No	No	N/A
	Q46	0	0	3	0	0	1	0	0	0	0	0	0	0	0	N/A
	Q47	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	N/A

*Comfort/Functionality Subscale. **Control/Privacy subscale.

Table L.4 Participants responses to Questions 24, 25, 28, 29, and 30 in the satisfaction questionnaire after installing the water feature.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Q24	Yes	No	No	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	No
Q24 Part 2	Payroll entry	N/A	N/A	Writing	HR record updating	N/A	N/A	N/A	Report Writing, Analysing data	Data entry, Telephone conversation	Reading, writing	Data entry, telephone conversation	Telephone Conversati on	N/A
Q25	Equally bothered by both	Hear only one speaker	Hear all speaker s	Equally bothered by both	Hear all speakers	Hear all speaker s	Equally bothered by both	Hear only one speaker	Hear only one speaker	Equally bothered by both	Equally bothered by both	Equally bothered by both	Equally bothered by both	Hear all speaker s
Q28	Yes	No	No	Yes	No	No	No	No	Yes	Yes	No	Yes	No	No
Q29	Payroll entry	N/A	N/A	Writing	N/A	N/A	N/A	N/A	Report writing	Data analysis	N/A	Data entry, telephone conversation	N/A	N/A
Q30	Shared office	Open- plan office	Shared office	Private office	Private office	Open- plan office	Shared office	Shared office	Open-plan office	Shared office	Shared office	Shared office	Other	Shared office